(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 April 2002 (11.04.2002)

PCT

(10) International Publication Number
WO 02/28876 A2

(51) International Patent Classification[7]: C07H 21/00

(21) International Application Number: PCT/JP01/08805

(22) International Filing Date: 5 October 2001 (05.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2000-306749 5 October 2000 (05.10.2000) JP

(71) Applicant *(for all designated States except US)*: **RIKEN** [JP/JP]; 2-1, Hirosawa, Wako-shi, Saitama 351-0198 (JP).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **HAYASHIZAKI, Yoshihide** [JP/JP]; 22-1-201, Inarimae, Tsukuba-shi, Ibaraki 305-0061 (JP).

(74) Agents: **SHIOZAWA, Hisao** et al.; 8th Floor, Kyobashi Nisshoku Bldg., 8-7, Kyobashi 1-chome, Chuo-ku, Tokyo 104-0031 (JP).

(81) Designated States *(national)*: CA, JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/28876 A2

(54) Title: OLIGONUCLEOTIDE LINKERS COMPRISING A VARIABLE COHESIVE PORTION AND METHOD FOR THE PREPARATION OF POLYNUCLEOTIDE LIBRARIES BY USING SAID LINKERS.

(57) Abstract: A linker or population of linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable portion represented by formula (N)n, wherein N is A, C, G, T or U, or their derivatives, and n is an integer equal to or higher than 1. A linker-polynucleotide or a population of linker-polynucleotides comprising said linker or population of linkers and a target first strand polynucleotide bound to said linker. A method of preparing said linker or population of linkers and a method of preparing a linker-polynucleotide using said linker or population of linkers. Provided is a linker instead of G tailing, which can be utilized in a method of preparing a cDNA library, and a method of preparing a cDNA library using said linker.

DESCRIPTION

Oligonucleotide linkers comprising a variable cohesive portion and method for the preparation of polynucleotide libraries by using said linkers.

TECHNICAL FIELD

The present invention relates to a population of linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable portion and to a method for the preparation of polynucleotide libraries comprising the use of said population of linkers. Further, the invention relates to an improved linker as a marker for specific libraries.

BACKGROUND ART

Oligonucleotide linkers and primers have been used in the prior art for priming, binding or annealing single strand polynucleotide and allowing the synthesis of the second polynucleotide complementary strand.

Carninci et al., 1996, Genomics, 37, 327-336; Carninci et al., 1997, DNA Research 4:61-66; Carninci et al., 1998, Proc.Natl.Acad.Sci USA, 95:520-4; Carninci and Hayashizaki, 1999, Methods Enzymol. 303:19-44, disclose method for the preparation of cDNA libraries. According to these protocols, a mRNA/cDNA hybrid is prepared and full-coding/full-length cDNAs are selected by mean of the Cap trapper technology, then

each single strand cDNA is ligated with G-tail and the cDNA second strand is synthesized.

However, the G-tailing methodology shows several drawbacks, for example, in the sequencing efficiency and translation efficiency when cDNAs clones are used for protein expression.

G-tailing is performed by terminal addition of dGTP using terminal deoxynucleotidyl transferase. However, the number of G added is difficult to control and it is variable, generally between 10 and 30. A long G-tail has the drawback of impairing a long read sequencing and lowering the sequencing efficiency, whilst a short G-tail has the drawback of providing a low efficient priming, with the consequence of lose of sample, and necessity of repreparing it.

During sequencing reaction, long G-stretches (long G-tail) interact with surrounding sequences and form very strong secondary structures. This may be problematic in case of interactions with 5' UTRs that are typically GC rich. In fact, a typical cDNA has 60% GC content in the 5'-UTR that is considered to act as a regulatory region. Similar problems were also observed in cloning vectors having GC rich region containing a *Sfi* I or *Not* I restriction sites next to cloning site.

Further, terminal deoxynucleotidyl transferase used for tailing reaction requires heavy metals, for example, $MnCl_2$ or $CoCl_2$. However, these heavy metals have sometimes caused degradation of cDNAs and decreased long strand,

full-coding/full-length cDNA production rate.

The purpose of the present invention is to solve the several problems in the prior art and provide a novel and efficient method for the preparation of cDNA libraries.

More specifically, the purpose of the present invention is to provide a novel linker instead of G-tailing, which can be utilized in a method for the preparation of cDNA libraries and to provide a method for the preparation of cDNA libraries using said linker.

DESCRIPTION OF INVENTION

The present invention solves the above-mentioned problems by providing a linker comprising an oligonucleotide fixed portion and an oligonucleotide variable portion, the variable portion being represented as Formula (N)n wherein N is A, C, G, T or U, or their derivatives, and n is an integer equal to or higher than 1. When n is an integer equal to or higher than 2, the nucleotides (N) of the variable portion may be the same or different.

The variable portion is preferably prepared at random.

The linker according to the invention may be a single or a double strand linker.

The present invention further relates to a population of linkers comprising the two or more of the linkers of the present invention.

According to an embodiment of the present invention, it is provided a linker or

population of linkers according to the present invention, which is prepared by;

(a) synthesizing a first oligonucleotide single strand comprising an oligonucleotide single strand fixed portion and an oligonucleotide single strand variable portion,

(b) synthesizing a second oligonucleotide single strand comprising an oligonucleotide single strand fixed portion complementary to the first oligonucleotide single strand fixed portion (a), and

(c) annealing the first oligonucleotide strand (a) to the second oligonucleotide strand (b) so that the variable portion protrudes outside the double strand fixed linker portion.

The present invention further relates to linker-polynucleotide product or population of linker-polynucleotide products comprising the linker or population of linkers according to the present invention and the target first strand polynucleotide bound to the linker.

The present invention still further relates to a vector comprising the linker-polynucleotide according to the present invention.

In addition, the present invention relates to a method for preparing the linker or population of linkers according to the present invention, which comprises the steps of:

(a) synthesizing a first oligonucleotide single strand comprising an oligonucleotide single strand fixed portion and an oligonucleotide single strand variable

portion,

(b) synthesizing a second oligonucleotide single strand comprising an oligonucleotide single strand fixed portion complementary to the first oligonucleotide single strand fixed portion (a), and,

(c) annealing the first oligonucleotide strand (a) to the second oligonucleotide strand (b), so that the variable portion protrudes outside the double strand fixed portion.

The present invention is further directed to the following methods:

(1) a method of binding a target single strand polynucleotide to a linker comprising:

i) the preparation of the linker according to the present invention; and

ii) the step of annealing the variable portion of said linker to the target single strand polynucleotide;

(2) a method of binding a target single strand polynucleotide to a linker comprising:

i) the preparation of the linker according to the present invention; and

ii) the step of annealing the variable portion of one (first) strand of said linker to the target single strand polynucleotide and ligating the fixed portion of the other (second) strand of said linker to the target single strand polynucleotide;

(3) a method of binding a target single strand polynucleotide or a population of the polynucleotides to a population of linkers comprising:

i) the preparation of the population of the linker according to the present invention; and

ii) the step of annealing the variable portion of said population of linkers to a population of the target single strand polynucleotides;

(4) a method of binding a target single strand polynucleotide or a population of the target single strand polynucleotides to a population of linkers comprising:

i) the preparation of the population of linkers according to the present invention; and

ii) the step of annealing the variable portion of the first strand of said population of the linkers to the target single strand polynucleotide or the population of the polynucleotides and ligating the fixed portion of the second strand of the population of the linker to the target single strand polynucleotide or the population of the polynucleotides;

(5) a method of preparing a linker-polynucleotide product comprising a linker and a double strand polynucleotide, comprising the steps of

i) annealing the variable portion of the linker according to the present invention to the target first strand polynucleotide, and

ii) synthesizing the second strand polynucleotide complementary to the target single strand polynucleotide;

(6) a method of preparing a linker-polynucleotide product comprising a linker and a double strand polynucleotide, comprising the steps of:

i) annealing the variable portion of the first strand of the linker according to the present invention to a target single strand polynucleotide and ligating the target single strand polynucleotide to the fixed portion of the second strand of the linker, and

ii) synthesizing the second single strand polynucleotide complementary to said

target single strand polynucleotide;

(7) a method of preparing a linker-polynucleotide product comprising a linker or a population of linkers and a population of double strand polynucleotides, comprising the steps of:

i) annealing the variable portion of the linker or a population of the linkers according to the present invention to a target single strand polynucleotide or a population of the target single strand polynucleotides, and

ii) synthesizing the second strand polynucleotide complementary to said target single strand polynucleotide or a population thereof;

(8) a method of preparing a linker-polynucleotide product comprising a linker or a population of linkers and a population of double strand polynucleotides, comprising the steps of:

i) annealing the variable portion of the first strand of the linker or a population of the linkers according to the present invention to a target single strand polynucleotide or a population of the target single strand polynucleotides,

ii) ligating the target single strand polynucleotide or the population of the target single strand polynucleotides to the fixed portion of the second strand of the linker or the population of the linkers, and

iii) synthesizing the second single strand polynucleotide(s) complementary to said target single strand polynucleotide(s);

(9) a method of marking a polynucleotide library and distinguishing said library, which comprises the step of providing a population of linkers comprising a fixed portion and a variable portion (wherein the fixed portion comprises at least one marker indicating

the defined tissue or species), and selecting and separating said library by said defined marker;

(10) a method of binding a linker or population of linkers to mRNA, which comprises the steps of:

(a) treating mRNA with phosphatase and removing phosphate groups from uncapped mRNA,

(b) treating a product of step (a) pyrophosphatase, which removes the CAP structure from capped mRNA, and

(c) adding an RNA ligase in the presence of the linker according to the present invention;

(11) a method of preparing a linker-polynucleotide product , which comprises the steps of:

(a) treating mRNA with phosphatase and removing phosphate groups from uncapped mRNA,

(b) treating a product of step (a) pyrophosphatase, which removes the CAP structure from capped mRNA,

(c) adding an RNA ligase in the presence of the linker according to the present invention, and

(d) adding an oligo dT and synthesizing a polynucleotide complementary to the complete sequence of said mRNA;

(12) a method of binding a linker and population of linkers according to the present invention to a target single strand polynucleotide or population of polynucleotides comprising the addition of RNA ligase;

(13) a method of preparing DNA/RNA hybrids, which comprises the steps of:

i) providing a full-length/coding or long poly-A mRNAs,

ii) ligating and annealing said mRNAs to the linker according to any one of claims 1 to 41, the linker comprising a first restriction enzyme site,

iii) annealing oligo dT-primers comprising the second restriction enzyme site, to the mRNA,

iv) synthesizing cDNA strands,

v) isolating the hybrids by using restriction enzymes which recognize the two specific restriction enzyme sites introduced, and

vi) cloning;

(14) a method of preparing a linker-polynucleotide product comprising a linker and a single strand polynucleotide, comprising annealing the variable portion of the linker according to the present invention to the target first strand polynucleotide.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows an example of the procedure for preparation of full-length cDNA library using, as an example, a population of linkers comprising the variable portion GNNNNN.

PolyA+RNA is transcribed (A) and subsequently oxidized and attached to biotin. After RNase I treatment (B), only full-length cDNA has biotin and trapped with avidin coated magnetic beads (C).

Figure 2 shows an example of the procedure for preparation of full-length cDNA library using, as an example, a population of linkers comprising the variable portion GNNNNN ,as a continuation of Fiure 1.

The cDNA is released from beads by alkaline treatment and recovered (D), and the linker is ligated (E). Shown is GN_5 linker. In the case of N6 linker, the variable portion is NNNNNN instead of GNNNNN. The second strand cDNA is synthesized (F) and digested with a restriction enzyme (G) and ligated into lambda phage vector (H) as well as packaged (I).

Figure 3 shows the result of the ligation between the tested cDNA and the linker.

The tested first strand cDNA prepared from 5 μg of 7.5kb poly(A)-tailed RNA (LifeTechnologies) was used as a starting material. The linker (GN_5-, Lanes 1 to 3; N_6, Lanes 4 to 6) was annealed to ligate with 50ng of 7.5kb tested cDNA. Subsequently, 10ng of linker-binding material sample was used for synthesizing the second strand cDNA, and then subjected to 0.8% alkali gel electrophoresis.

Linkers were used in different amounts: 200ng for Lanes 1 and 4; 500ng for Lanes 2 and 5, and 2 μg for Lanes 3 and 6). As a control, cDNA without linkers was used as a template of the second strand synthesis (Lane 7). Lane 8 comprises the first strand cDNA without linkers. Lane 9 comprises λ/HindDIII size marker.

Figure 4 shows the result of examining the ratio between a linker ligation cDNA of intercellular cDNA and a linker, using linkers with various molar ratios.

A linker combined with 2 μg of N_6/GN_5 (N_6:GN_5=1:4) was ligated with various quantities of cDNAs (1000ng for Lane 1, 500ng for Lane 2 and 200ng for Lane 3).

They were supplied for the second strand cDNA synthesis and analyzed by 0.8% gel electrophoresis. Lane 4 comprises markers.

The indivisual lane on the left side of the Figure refers to the first strand cDNA.

Figure 5 shows a sequencing chart of a cDNA sequence having G-tail described in the prior art.

In the presence of a repetition of C in the second strand cDNA (introduced with the G-tail in the first strand), the efficiency of sequencing dropped down as shown in the chart.

Figure 6 shows a sequencing chart of a cDNA sequence ligated with N_6/GN_6 linker mixture (proportion 1:4).The sequenced clone D05_042_2-5F-ab2 in Figure 6 corresponds to the sample 2.05 in Table 1.

Figure 7 shows a sequencing chart of a cDNA sequence ligated with GN_5 linker. The sequenced clone G07_052_3-7F.ab1 in Figure 7 corresponds to the sample 3.07 in Table 1.

Figure 8 schematially shows a loop bias and the possible solutions. In Fig.8(A), the bias are explained. One of the end of the fixed or constant portion of the single strand linker, after removal of the other single strand from the double strand linker, may interact with the single strand cDNA to form a loop and block the following synthesis of the second strand cDNA.

As shown in Fig.8(B), NH_2 as a protecting group is bound to 3' end in the case of one end of the constant or fixed lower strand. There is no possibility of forming a loop, as well as the second strand cDNA synthesis is not inhibited.

In Fig.8(C), 3' end of the fixed or nonvariable second strand (lower strand in the figure) and 5' end of the fixed or nonvariable first strand (upper strand in the figure) of a linker are bound together to form a loop. This prevents the possiblity of forming a loop with a single strand cDNA, as well as the second strand cDNA synthesis is not inhibited

DETAILED DESCRIPTION OF THE INVENTION

The present invention solves the above-mentioned problems in the prior art by providing a linker and a population of linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable portion. Such a linker and a population of linkers can bind to an end of a target single strand polynucleotide or a population of target polynucleotides, as well as allow the second polynucleotide strand synthesis.

According to the embodiment of the present invention, provided are a linker and a population of linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable portion. The fixed portion is preferably an oligonucleotide portion which is nonvariable in any linker of a population of linkers. The linker fixed portion can be a single strand or double strand oligonucleotide, preferably, a double strand oligonucleotide. The fixed portion preferably comprises at least one of the followings: a restriction site, a recombinational site, a polymerase promoter site, a marker or a tag.

The variable portion is preferably synthesized at random. The linker of the resulting population of linkers has a nonvariable portion, preferably the common portion among the population and a variable portion which is different for each linker among the population.

The population of linkers according to the present invention may comprise, as a variable portion, one or more linkers having an oligo sequence specific for the 3' or 5' end of a target single strand polynucleotide. Said oligo sequence is specifically selected in order to bind and isolate one or more specific target polynucleotides among a population of target single strand polynucleotides.

The linker variable portion according to the present invention may also work as a primer in the synthesis of the second strand polynucleotide, as for example a long strand full-cording or full-length cDNA.

Said randomly variable single strand oligonucleotide portion can comprise any kind of nucleotide. Preferably, the variable portion has the formula (N)n, wherein N is A, C, G, T or U, or their derivatives, and n is equal to or higher than 1. When the integer n is equal or higher than 2, the nucleotides of the variable portion may be the same or different from each other.

The integer n is advantageously is from 1 to 10, preferably from 4 to 8, more preferably n is 5 or 6.

As one method, in the variable portion the first to third nucleotides (beginning counting from the side of the fixed portion to the free end as shown in Fig.2) of (N)n can be G. A mixture having different strand length (that is, the length of n), and presence or absence of G is also within the object of the present invention. Preferably, it is a mixture of N_6/GN_5 with different proportion and the proportion is preferably 1:4.

The present invention further relates to a linker-polynucleotide product comprising the linker according to the present invention and a single strand or double strand polynucleotide annealed and/or ligated to the variable portion of said linker, and to a vector comprising said linker-polynucleotide product .

Preferably, said single or double strand polynucleotide is a long strand, full-coding/full-length cDNA.

Accordingly, the present invention discloses a method for the preparation of linker-polynucleotide products or polynucleotide libraries comprising the linkers according to the present invention annealed and/or ligated to target single strand polynucleotides.

The present invention further relates to a method for the preparation of polynucleotide products or libraries comprising the linkers according to the invention annealed/ligated to a double strand polynucleotides, preferably, to a method for the preparation of long strand, full-coding/length cDNA libraries.

The present invention further relates to a method for marking polynucleotide libraries by providing a linker population according to the invention comprising a marker in the fixed portion. This marking system allows to distinguish and recognize libraries of different species (for instance, human, mouse, Drosophila, rice, and the like) and it can be used for distinguishing libraries of different tissues (for instance, liver, brain, lungs, and the like) for each species.

According to the present invention, provided is a method for binding a linker and a linker population described below and said linker or linker population with a target single strand polynucleotide or a population of target single strand polynucleotides, comprising the steps of:

i) preparing a linker or a population of linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable single strand portion;

ii) annealing a target single strand polynucleotide(s) to the variable portion(s) of said linker.

Hereinafter, in some cases, the present invention, for simplicity, will be described for a population of linkers (also indicated as a population of linkers or simply linkers) as well as to a method for binding said population to target polynucleotides and to a method for synthesizing double strand polynucleotides. However, it is clear that the present invention also includes the indivisual linkers comprised in said population of linkers and a method comprising the use of such indivisual linker.

The fixed portion of each linker comprised in the population of linkers according to the present invention can be a single or double strand oligonucleotide.

Preferably, the fixed portion, as well as the linker, is a double oligonucleotide, accordingly, the method comprises the steps of:

i) preparing a linker or a population of linkers comprising an oligonucleotide double strand fixed portion and oligonucleotide variable single strand portion, wherein said variable portion is protruding outside said double strand fixed portion (therefore, the variable portion forms a cohesive protruding end);

ii) annealing a population of target single strand polynucleotides to the variable portions of the population of linkers, and preferably ligating the annealed end of said target single strand polynucleotides to the adjacent fixed portions of linkers (see steps (F) to (G) in Figure 2).

Said linker or linkers can be prepared with any methodology known in the prior art, for example, with one oligonucleotide strand having the direction 5'-3', comprising a fixed nucleic acid sequence at the 5' portion and a variable nucleic acid sequence end at the 3' portion. Then, the other oligonucleotide strand comprising a fixed nucleic acid sequence is prepared. Finally, the fixed portion of one strand and the fixed portion of the other strand are annealed so that the variable portion of one strand protrudes outside the double strand fixed portion. As a matter of course, the linker can also be prepared with inverted order of steps and with other methodologies.

When the linker according to the present invention is a double strand linker, for the purpose of the present application, the strand comprising the fixed portion and the variable portion is also referred to as "the first strand", while the other strand complementary to the fixed portion of the first strand is referred to as "second strand".

The present invention further relates to a linker comprising only one oligonucleotide strand comprising a fixed portion and a variable portion (that is only comprising the "first strand" and not comprising the "second strand") . If the linker has the direction 5'-3', the variable portion of the end of this single strand oligonucleotide anneales to the 3' end of a target single strand polynucleotide. Then, the second strand polynucleotide, complementary to this target single strand polynucleotide, is synthesized.

One or more linkers can also be prepared in such a way that the strand comprising a variable portion has a direction 3'-5'. As a result, the variable portion is positioned at the 5' end. The 5' end variable portion of the linker thus prepared anneals to the 5' end of the target polynucleotide. If the other strand is present, its lower strand can be ligated to the 5' end of the target polynucleotide.

The present invention also provides a method for preparing linker-polynucleotide products comprising the linker or the population of linkers according to the present invention and a double strand polynucleotide, comprising the steps of:

i) preparing a population of linkers comprising an oligonucleotide double strand fixed portion and an oligonucleotide variable single strand portion, wherein said variable portion is protruding outside the double strand fixed portion (therefore, the variable portions form cohesive protruding ends);

ii) annealing a population of the target single strand polynucleotides to the variable portions of said population of linkers, and ligating said population of target single strand polynucleotides to the adjacent (second strand) fixed portions of said linkers; and

iii) synthesizing second single strand polynucleotides, complementary to the target single strand, by using the variable portions as primers.

The polynucleotide sequence can also be prepared using only one strand of the linker according to the present invention. In this case, comprised are the steps of:

i) preparing a population of single strand linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable single strand portion;

ii) annealing a population of the target first strand polynucleotides to a variable portion of the population of linkers;

iii) synthesizing second single strand polynucleotides, complementary to the target first strand, by using the variable portions of linkers as primers.

The fixed oligonucleotide portion of the linker or population of linkers can be any oligonucleotide sequence. This fixed sequence is preferably a nonvariable portion and it is therefore the common for all the linkers of the same population. This fixed

portions can also comprise oligo sequences consisting of one or more groups and therefore the fixed portions, in this case, can show some differences among the same population. However, since these oligo sequences consisting of one or more optional groups will not change the general structure of the fixed portions, for the purpose of the present invention, the fixed portion, comprising or not comprising the variable oligo sequences consisting of one or more groups will be, for simplcity, indicated as fixed portion.

The fixed portion can be a nonvariable portion even in the population of linkers. That is, it can be the same one for any linkers comprised in the population.

The fixed portion can be any oligonucleotide sequence (DNA or RNA), and it is preferably the same or almost the same for the linker or the population of linkers used in the specific experiments or for the specific library.

The linker fixed portion can therefore be intended both as a single or double strand oligonucleotide, preferably, it is a double strand oligonucleotide. In this case, the single strand variable portion constitutes a protruding end. The variable portion can act as a primer in the second strand polynucleotide synthesizing process.

The fixed (or nonvariable) portion preferably comprises one or more restriction site, homologous recombinational site, polymerase promoter site, a marker and/or a tag. Preferable restriction sites are, for example, BamHI, XhoI, SstI, SalI or NotI and others,

for example those disclosed in Hyone-myong Eun, Chapter "Restriction endonucleases and modification methylases" .

Examples of homologous recombinational sites are attB, Gateway™ (Life Technologies), Cre-lox (Qinghua Liu, et al., 1998, Current Biology, 8:1300-1309) Flp/FRT (J. Wild, et al, 1996, Gene, 179:181-188).

Further, as for the polymerase promoter site, it can be a RNA polymerase promoter site, for example one of those cited in Hyone-Myong Eun, page 521. Preferably, it can be T3, T7,SP6, K11 and/or BA14 RNA polymerase promoter site.

A marker can be any sequence or sequences of nucleotides, for example a sequence specific for a particular tissue or species.

As a tag, any group or molecule able to be bound to an end of the fixed portion of one strand or the other strand can be used. In fact, when a single strand of the linker is removed, for example by increasing temperature, the end of the other single strand could form a loop with the target single strand polynucleotide (Fig. 8). Preferably, a protecting group is bound to the 3' end of the strand consisting of only this fixed portion in order to avoid that the end of the strand consisting of only the fixed portion forms a loop with the target single strand polynucleotide mainly ligated to this strand, as well as inhibits the synthesis of the second strand polynucleotide (Figure. 8A). Therefore, any

group that does not have a 3'-OH and cannot be ligated nor extended by DNA polymerases, can be used for the purpose of the present invention.

As a protecting group, for example ddNTPs can be used. Preferably, a NH_2 group is also used as a protecting group (Figure 8, B).

As further particular solution, in order to avoid the problem of the loop bias, both ends can be bounded together so that the ends of both strands of the fixed portion of the linker positioned opposite to a variable portion form a loop (Figure 8, C). With this solution, the ends of the fixed portion cannot form a loop with the target single strand polynucleotide, and the synthesis of the second strand polynucleotide is not inhibited.

The oligonucleotide variable portion of the linker or the population of linkers is, preferably, randomly synthesized. Accordingly, in a population of linkers, the variable portion of any linker is preferably synthesized at random and the sequence of the variable portion and/or the number of base comprised in each linker differs each other. A population of linkers, therefore, comprises protruding ends having a high number of different sequences. Such a population of linkers comprises a high variation of random protruding ends. These recognize, anneal and/or ligate to the complementary ends of a population of target single strand polynucleotides. That is, this is a population of full-length cDNAs forming polynucleotide sequences comprising the linker and a target single strand polynucleotide (see Figures 1 and 2).

The present invention therefore also relates to a population of linkers comprising at least two linkers prepared according to the present invention. Preferably, the invention relates to a population of linkers comprising at least two subpopulations of linkers.

The population of linker can be that one in which the fixed portions comprised in all the linkers are an oligonucleotide portion having the same sequence.The population of linkers may also comprises two or more subpopulations of linkers, wherein one subpopulation of linkers comprises linkers in which the fixed portion is an oligonucleotide portion having the same sequence, and other subpopulations of which differ each other in the fixed portion sequence.

Preferably, in the population or subpopulation of linkers the variable portions of the linkers are synthesized at random. Preferably, in the population or subpopulation, the sequence of the variable portions of the linkers are different from each other.

The variable portion can also be a specific oligonucleotide sequence complementary for an end (preferably 3' end) of a target single strand polynucleotide.

Preferably, the linker population of the present invention comprises among the variable portions, one or more specifically determined portions able to recognize and anneal end(s) of specific target polynucleotides which are intended to select from the population of target polynucleotides.

The end of a target single strand polynucleotide anneals to the protruding variable end of the linker. When a population of linkers is added to a population of target single strand poynucleotides, the variable portions (protruding ends), preferably randomly synthesized, in the linkers recognize and anneal to the ends of the population of target single strand polynucleotides.

Preferably, the linkers according to the invention are double strand oligonucleotides comprising a fixed portion (preferably a nonvariable portion for all the linkers of a population) and a variable portion, which is different for any linker of the population. According to the first embodiment, the 3' end of the target single strand polynucleotide anneals to the 3' end of protruding end of the variable portion and ligates to the 5' end off the fixed portion of the other strand adjacent to the 3' end of said target single strand polynucleotide.

The linker can also be constituted, according to the second embodiment, by the fixed portions of one strand having the direction 3'-5' and the other strand. In this case, the 5' end of the target single strand polynucleotide anneals and ligates to this variable portion of the linker.

The variable single strand oligonucleotide portion of the linker can comprises any kind on nucleic acid. Preferably, said variable portion has the formula (N)n, wherein N is A, C, G, T or U or their derivatives and n is equal to or higher than 1, and if n is an integer equal to or higher than 2, the nucleotides (N) of the variable portion may

be the same or differ from each other. Preferably, $1 \leqq n \leqq 10$ and more preferably $4 \leqq n \leqq 8$. As a particular preferred linker, n is 5 or 6, that is, N_6 or N_5.

The first, second and/or third N, closest to the fixed portion (that is, the nucleotides of the variable portion coming from the 5' end of the linker in case of the first embodiment) can also be a G, according to the formula $(G)_m(N)_{n-m}$, wherein m = 1 to 3. Preferably, the linker variable portion can be GN_4, GN_5, G_2N_3, G_2N_4, G_3N_2, G_3N_3, N_5, N_6 or a mixer thereof.

More specifically, the linker population according to the present invention is a mixture of $(N)_n$ linker and $(G)_m(N)_{n-m}$, preferably N_6/GN_5, N_6/G_2N_4 or N_6/G_3N_3 having different proportion. The proportion of the N_6/GN_5 mixture linker can be 0:1- 1:0, preferably 1:3-1:5, more preferably 1:4. The ligation can be realized with any ligation method known in the prior art, preferably by using a DNA ligase, more preferably a T4 DNA ligase or E. coli DNA ligase (see for example, Hyone-Mong Eun, under the chapter "Ligases") or using a RNA ligase (Maruyama et al. 1995).

Preferably, the ligation reaction according to the present invention comprises the addition of ligase stimulating agents. Preferably, as a ligase stimulating agent, PEG (Polyethylene glycol), preferably at 6000-8000 molecular weight, is used.

After the annealing and/or ligation step, the variable linker portion (that is, the protruding or free 3' end of the linker of the first embodiment) can act as a primer for the

synthesis of the second strand polynucleotide, forming a polynucleotide sequence comprising the linker according to the invention and a double strand polynucleotide.

The ligation of the linker according to the invention to a target single strand polynucleotide can also be realized by using the oligo-capping technique (K.Maruyama et al., 1995, Gene, 138:171-174; and S.Kato et al., 1995, Gene, 150:243-250). The oligo-capping essentially comprises the following steps: i) mRNAs extracted from cells are treated with a phosphatase enzyme, preferably bacterial alkaline phosphatase for removing phosphates from non full-length mRNAs (that is, forming 5' ends of uncapped RNA having an hydroxyl at the 5' end, but not removing the CAP structure from the capped full-length RNA); ii) the mixture obtained in i) is treated with a pyrophosphatase, preferably tobacco acid pyrophosphatase (TAP), which removes the CAP structure from full-length mRNAs and leaves the full-length 5' ends with a phosphate group; iii) the full-length mRNA having a phosphate group at the 5' end is ligated to a specific RNA or a DNA adapter with a RNA ligase; and iv) an oligo dT is added and the complementary strand is synthesized.

The method for binding the linker according to the present invention to a target polynucleotide and/or the method of preparing a polynucleotide sequence according to the present invention can also be performed using, as ligation step, a modified oligo-capping method as follow.

Accordingly, a target single strand polynucleotide (which can be RNA, mRNA prepared as described as the oligo-capping method, or cDNA) can be ligated to the linker according to the present invention in presence of a ligase.

In particular, when the linker of the present invention is a double strand linker, an end of the target single strand polynucleotide ligates to the second strand (which consists of the only fixed portion) and anneals to the variable portion of the first portion of the linker.

As an another possibility, the target polynucleotide ligates to the variable portion of the linker (which can be both a single or a double strand). In both cases, an oligo dT is added and a complementary polynucleotide, preferably cDNA, is synthesized.

The use of RNA ligase is not limited to RNA or mRNA as above described, but can also be used to ligate DNA.

The ligation method using RNA ligase in order to bind the linker according to the present invention and a target single strand polynucleotide, can therefore bind:
i) a single strand DNA to a single strand DNA; ii) a single strand RNA to a single strand RNA; and iii) a single strand DNA to a single strand RNA or a single strand RNA to a single strand DNA.

According to an embodiment, the polynucleotide is a long strand, full-coding/full-length mRNA and the linker is DNA (but can also be RNA) and comprises a first restriction enzyme site.

Accordingly, it is provided a method for the preparation of a single or double strand cDNA comprising the steps of:

(I) providing a long strand, full-coding or full-length mRNA comprising a poly-A;

(II) providing a double strand linker comprising a first restriction enzyme site;

(III) ligating the 5' end of the mRNA (by using a ligase, for instance RNA ligase) to the fixed portion of the second strand of the linker and annealing the 5' end to variable portion of the first strand of the linker;

(IV) providing an oligo dT-primer comprising a second restriction enzyme site, and annealing the oligo dT-primer to the poly-A of the mRNA;

(V) synthesizing the cDNA by addition of a reverse transcriptase and NTPs; during this step, the new synthesized cDNA displaces the linker first strand (that one comprising the fixed and the variable portion);

(VI) removing the mRNA and obtaining a single strand cDNA.

Further, a primer can be added to the 3' end of the cDNA, and in presence of a polymerase a complementary DNA is synthesized forming a double strand cDNA. The double strand formed therefore comprises a first restriction enzyme site at one end and a second restriction enzyme site at the other end.

The removal of mRNA at step VI) may be performed by addition of a RNase H or other enzyme which cuts RNA in fragments and remove them, or by addition of alkali (for instance NaOH) according to the methodology known in the art (Sambrook et al, 1989).

The double strand polynucleotide sequence is then cleaved at the first and second restriction enzyme sites, specifically introduced with the linkers, by using the specific restriction enzymes, resulting in forming protruding ends. The double strand polynucleotide with protruding ends is then inserted in a plasmid or phage expression vector or in a sequencing vector (for example, as described in, for example, Sambrook et al, 1989, Molecular Cloning, Cold Spring Harbor Laboratory; Invitrogen Catalog 1999; Stragene Catalog 1999, etc.). The double strand polynucleotide can also be cloned by site-specific recombination (for example attB-attP) or by blunt-ends methodology (Sambrook et al., 1989).

Examples of phage vectors are lambda-ZAP, lambda-Dash (Stratagene).

The invention, therefore, is also related, but not limited, to a phage or plasmid expression or sequencing vector comprising the polynucleotide sequence according to the present invention.

The single or double strand polynucleotide according to the present invention is RNA or DNA, or also a DNA/RNA hybrid. Preferably, a long strand full-coding and/or

full-length cDNA. The 3' end of said long strand full-coding/full-length cDNA corresponds to the 5' Cap end of mRNA.

For the purposes of the present invention, with the wording full-length cDNA is intended a cDNA comprising the 5' and 3' UTR sequences and the oligo dT-primer (that is, complementary to a mRNA comprising the poly-A). It may also comprises additional sequences for cloning, such as restriction enzyme sites. With a full-coding cDNA, a cDNA sequence comprises at least the start and stop codon. And with long strand cDNA, it is understood a cDNA sequence which is almost full-coding/full-length, lacking of one or few nucleotides at the 3' end (corresponding to the 5' end of mRNA) or at the 5' end if considering a cDNA strand complementary to the cDNA complementary to the mRNA (that is, having the same direction of the gene). Such a stop of synthesis reaction during cDNA synthesis may be caused by the formation of secondary structure of the mRNA, for example, the level of the Cap structure. However, also fragments of genes, nucleotides, cDNAs, RNA or mRNA are not excluded from the purpose of the application of the present invention.

A DNA/RNA hybrid can be prepared by:

(I) providing a long strand, full-coding or full-length mRNA comprising a poly-A;

(II) providing a double strand linker comprising a first restriction enzyme site;

(III) ligating the 5' end of the mRNA (by using a ligase, for instance RNA ligase) to the fixed portion of the second strand of the linker and annealing the 5' end to variable portion of the first strand of the linker;

(IV) providing an oligo dT-primer comprising a second restriction enzyme site, and annealing the oligo dT-primer to the poly-A of the mRNA;

(V) synthesizing the cDNA by addition of a reverse transcriptase and NTPs; during this step, the new synthesized cDNA displaces the linker first strand (that one comprising the fixed and the variable portion);

(VI) adding an oligonucleotide complementary to the second restriction enzyme site of the oligo dT-primer and ligated this oligonucleotide to the poly-A; an hybrid double strand polynucleotide is then formed.

The hydrid double strand polynucleotide can be cleaved by specific restriction enzymes as above said and inserted into a vector as above.

The target single strand polynucleotide, which anneals the variable linker portion and/or ligates to the adjacent fixed linker portion of the linker, can be prepared with any technique known in the prior art.

Preferably, the long strand full-coding/full-length single strand cDNAs are prepared according the technique of 5' mRNA Cap trapping, disclosed in Carninci et al., 1996, Genomics, 37, 327-336; Carninci et al., 1997, DNA Research 4:61-66; Carninci et al., 1998, Proc.Natl.Acad.Sci USA, 95:520-4; Carninci and Hayashizaki, 1999, Methods Enzymol. 303:19-44.

Preferably, all the steps described in the above prior art documents are followed, with the exception that instead of the G-tailing step, the population of linkers according to the invention is provided.

Preferably, the Cap-trapping method described in Figures 1 and 2 is used, however, the target single strand polynucleotide is not limited to that prepared with this technology. For example, other methods of isolation of first strand cDNA such as that described in Edery et al., 1995, Mol Cell Biol, 15:3363-71 or the oligo-capping method (K.Maruyama et al., 1995, Gene, 138:171-174; and S.Kato et al., 1995, Gene, 150:243-250) can also be used.

The target single strand polynucleotides according to the present invention can also be normalized and/or subtracted (for example, Soares et al., 1994, Proc. Natl. Acad. Sci. 91:9228-9232 and Bonaldo et al., 1996, 6:791-806). The recovered normalized and/or subtracted polynucleotides, preferably cDNAs, more preferably long strand full-coding/full-length cDNAs, are preferably prepared according to the Cap-trapping technology, and then ligated to the population of linkers according to the present invention, or said isolated cDNAs are first annealed and/or ligated to the population of linkers of the present invention, and then normalized and/or subtracted.

The target single strand polynucleotide may show a bias due to the formation of a loop or hairpin-loop. For example, the 3' end of a synthesized intercellular cDNA

may form a loop with an internal portion of itself, preventing the following annealing and ligation with the linker according to the present invention.

In order to solve this problem, the target single strand polynucleotide is optionally subjected to high temperature, from 25°C to up the boiling point of the solution (about 100°C), preferably at 65°C, and then cooled down, preferably in ice, before annealing and/or ligation with the linker according to the present invention.

As a modified method, the secondary structure can be deleted with chemical agents, such as solutions consisting of NaOH (for example 0.1 N), formamide 50-99% and Urea 6-8M or similar agents known to delete/reduce the secondary structure of nucleic acids or denature the double strand nucleic acids. In this case, such agents must be removed, usually by ethanol precipitation, prior to subsequent enzymatic reactions.

As a further modified method, the target polynucleotide annealed and/or ligated to the linker can be subjected to high temperature (hot start) in order to remove possibility of hairpin-loop formation. The temperature range is from 25°C up to the boiling point of the solution (about 100°C), preferably 65°C.

However, the increase of temperature may remove one strand of the linker (that is, the strand comprising the fixed portion and the variable portion), therefore later the same strand linker or any primer can be added to the other strand of the linker and the polynucleotide sequence comprising the target single strand polynucleotide. There is a

possibility of the formation of hairpin-loop, but it can be avoided using the solutions described in Figure 8, B and C.

Using the method according to the invention, the annealing and ligation step result to be very efficient, so that the following cloning step allows the preparation of high-titer libraries without PCR amplification.

The method according to the invention allows the preparation of libraries more advantageously compared to the method in the prior art and in particular to the method of G-tailing.

The G-tailing method in fact has a serious drawback during the sequencing process, as shown in Fig. 5. The second strand cDNA comprises a repetition of C, complementary to the G-tail sequence (the length of which cannot be easily controlled and therefore may reach the length of 20-30 G). This excessive C strength repetition makes the sequencing process to stop, preventing the DNA sequencing.

The method using the linker according the invention does not have this drawback (even if the random variable portion comprises G, they are statistically within a small number) and can allow an efficient sequencing as described in Figures 6 and 7.

The clone of Figure 6 comprises a portion of a linker N6 (marked in the box of Fig.6) corresponding to nucleotide 12 to 49 of SEQ ID NO:3. The nucleotides 1 to 11

(included) were cleaved as shown at step G of Figure 2. The variable portion of linker of Fig. 6 is GGCGAA (as shown in the marked box).

The clone of Figure 7 comprises a portion of a linker GN5 (marked in the box of Fig. 7) corresponding to nucleotide 12 to nucleotide 49 of SEQ ID NO:1. The nucleotides 1 to 11 (included) of SEQ ID NO:1 were cleaved as shown at step G) of Figure 2. The variable portion of linker of Fig.7 is GGCGAA (as shown in the marked box).

Then, the long G-stretches of the G-tailing methodology may interact with surrounding sequences and form very strong secondary structures, and this phenomenon affects the efficiency of sequencing, transcription and translation. On the contrary, the linker according to the present invention does not have these drawbacks.

Further, terminal deoxynucleotidyl transferase used for G-tailing reaction requires the presence of heavy metals, like $MnCl_2$ or $CoCl_2$. These heavy metals cause degradation of cDNAs and decreased long strand full-coding/full-length cDNA content rate. Also this problem is solved using the linker according to the present invention, which does not require heavy metals and can be performed at low temperature, for instance, 4-37°C, preferably 12-20°C, or preferably 16°C.

According to another embodiment of the present invention, the constant portion of the linker of the present invention can comprises a marker. For example, a specific

oligonucleotide sequence, a specific sequence or combination of sequences, easily recognizable.

The presence of this marker is very useful in order to distinguish and not to confuse libraries of different tissues (for instance, liver, brain, lungs, and the like) for the same or for different species, or libraries of different species (for instance, human, mouse, Drosophila melanogaster, rice, and the like).

In fact, when many kinds of libraries obtained from different tissues and/or species are constructed in the same laboratory and used for large-scale sequencing, there is the risk of confusing or contaminating the libraries or clones at any stage of colony picking, DNA preparation, sequencing determination, clones banking, re-arraying, etc.

Individual marking of cDNAs allows the preparation of different marked cDNAs from several tissues, allowing tissue expression profiling by sequencing 3' ends (complementary to the 5' mRNA end) of mixed cDNA libraries.

[Example]

The method and embodiments according to the present invention will now be illustrated with reference to the following examples.

Example 1

Linker evaluation using a Test cDNA

Linkers preparation

The population of linker oligonucleotides were purchased by Gibco-BRL Life technologies. The oligonucleotides were distinguished in one single strands (single upper strands) (indicated as A and C, comprising the variable portion) and the other single strands (single lower strands) (indicated as B). Then, one of A and C and the B were bound together in order to form two different population of double strands. The population of linkers A comprises linkers having a fixed portion oligonucleotide (in this case, bases 1-43 of SEQ ID NO:1) and a variable portion (GN_5), wherein the first base is G (that is, the base number 44) and the following bases NNNNN (bases from 45 to 49) different for each linker of the population and prepared at random.

The population of linkers C, comprises a constant portion oligonucleotide (bases 1-43 of SEQ ID NO:3) and the variable portion NNNNNN (bases 44 to 49) different for each linker of the population and prepared at random.

A) GN_5 A strand,

5'-AGAGAGAGAGCTCGAGCTCTATTTAGGTGACACTATAGAACCAGNNNNN-3' (SEQ ID NO:1);

B) B strand,

5'-TGGTTCTATAGTGTCACCTAAATAGAGCTCGAGCTCTCTCTCT-3' (SEQ ID NO:2);

The B strand was also phosphorylated at the 5' end when it was synthesized.

C) N_6 C strand,

5'-AGAGAGAGAGCTCGAGCTCTATTTAGGTGACACTATAGAACCANNNNNN-3' (SEQ ID NO:3).

For degenerate nucleotides, V stands for A, G or C and N stands for any nucleotide, according to the international convention and to the Patentin Standard 2.1 Manual.

These oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.) in order to remove contaminants that may have non-specific or lack annealing sites. Two populations of linkers named GN_5 and N_6 were prepared. The linker GN_5 was made by oligonucleotides A/B (SEQ ID NO:1/SEQ ID NO:2) and the linker N_6 was made by oligonucleotides C/B (SEQ ID NO:3/SEQ ID NO:2). They were prepared by mixing the oligonucleotides with NaCl (final concentration, 100 mM) and incubating at 65℃ for 5 min, 45℃ for 5 min, 37℃ for 10 min and 25℃ for 10 min.

The linkers prepared were then used for annealing to and ligating with single strand DNA(s).

Test cDNA

To establish the appropriate linker when preparing cDNA libraries, a test first strand cDNA has been generated from 5 μg 7.5-kb poly(A)-tailed RNA (Life Technologies) according to the method described in Carninci and Hayashizaki, 1999, except that CAP-Trapping was omitted. [α-^{32}P]dGTP was incorporated at the reverse-transcription step. The amount of the produced first strand cDNA was estimated according to the incorporation ratio of radioactivity. Then, 50 ng of the 7.5-kb cDNA and various amounts (200 ng to 2 μg) (see also description of Figure 3) of linker (N_6 or GN_5), prepared in the above step of Example 1, were combined together and ligated in 30 μL reaction volume. The reactions were performed by the incubation overnight at 10° C.

After ligation, to remove excess linkers, linker-bound single strand cDNA samples were incubated with 0.2 mg/mL proteinase K in 10 mM EDTA/0.2% SDS (reaction volume, 40 μL) at 45° C for 15 min. The reaction products were extracted by using phenol/chloroform 40 μL. Then, the phenol/chloroform mixture was treated with 60 μL column buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1 M NaCl, 0.1% SDS; pH 7.5) in order to extract the reaction product which was still remaining in the interface of the phenol/chloroform mixture. The reaction products extracted were loaded on gel-filtration columns Sephacryl™-300 (Amersham Pharmacia Biotech) and purified by centrifugation at 400 x g for 2 min. The eluted fractions (comprising the purified linker-first strand cDNA samples) were precipitated by using isopropanol.

The control sample (comprising the single strand Test cDNA but no linker) was used for the synthesis of the second strand cDNA. To assess the ability of the present method to support the synthesis of second-strand cDNA, 10 ng of the purified linker-ligated samples (Lanes 1-6 of Figure 3) and unligated 7.5-kb first-strand cDNA as a control (Lane 7 of Figure 3) were independently combined in a 10-μL reaction solution containing 1 μL 10X ExTaq™ buffer, 1 μL 2.5 mM dNTPs, 0.5 μL [α-^{32}P]dGTP, and 0.5 μL Ex-Taq™ (Takara). The obtained samples were incubated at 65° C for 5 min, 68° C for 30 min, and 72° C for 10 min and then analyzed by using alkaline gel electrophoresis.

The alkaline gel electrophoresis was performed by adding 5 μ1 of the samples to 1 μ1 of 6 x Alkaline dye (Sambrook, Molecular Cloning, 6.7, 6.12). Used were the electrophoresis gel contained 0.8 % of agarose, 50 mM NaOH and 5 mM EDTA and the buffer contained 50 mM NaOH and 5 mM EDTA (Sambrook, Molecular Cloning).

Results

Lanes 1-3 indicates the ligation of 50 ng between single strand Test cDNA and respectively 200 ng, 500 ng and 2 μg of linker GN_5.

Lanes 4-6 indicates the ligation between 50 ng of single strand Test cDNA and respectively 200 ng, 500 ng and 2 μg of linker N_6.

Lane 7 is the control. 10 ng of single first strand cDNA were added to 10-μL reaction solution comprising 1 μL 10X ExTaq™ buffer, 1 μL 2.5 mM dNTPs, 0.5 μL [α-^{32}P]dGTP, and 0.5 μL Ex-Taq™ (Takara). The sample was incubated at 65° C for

5 min, 68° C for 30 min, and 72° C for 10 min. The single strand cDNA was extended by forming a hairpin structure to form a second strand cDNA. This was detected at 15 kb in the alkaline gel electrophoresis.

Lane 8 comprises the first strand cDNA (23 ng) (without linker). This is detected at 7.5 kb.

Lane 9 represents the markers..

The electrophoresis of Figure 3 shows that in Lanes 1-6, the ligation was particularly efficient (spots at level of 7.5 kb) and the amount of no ligation was negligible (spots at level of 15 kb).

Example 2

Full-length cDNA library preparation and cDNA analysis

Linker preparation

Linkers were prepared as above described in Example 1.

Preparation of RNA

Slices of mouse liver tissue (0.5-1g) were homogenized in 10 ml of a suspension and extracted with 1 ml of 2M sodium acetate (pH 4.0) and the same amount of a mixture of phenol/chloroform (volume ratio 5:1). After the extraction, the same volume of isopropanol was added to the aqueous layer to precipitate RNA. This sample was incubated on ice for an hour and centrifuged at 4000 rpm for 15 minutes with cooling to collect the precipitates. The resulting precipitates were washed with 70% ethanol and dissolved in 8 ml of water. By adding 2 ml of 5M NaCl and 16 ml of an aqueous solution (pH 7.0) containing 1% CTAB (cetyltrimethylammonium bromide), 4M urea

and 50 mM Tris, RNA was precipitated and polysaccharides were removed (CTAB precipitate). After centrifugation at 4000 rpm for 15 minutes at room temperature, the RNA was dissolved in 4 ml of 7M guanidine-Cl. Then, two-fold amount of ethanol was added to the solution, incubated for an hour on ice and centrifuged at 4000 rpm for 15 minutes. The resulting precipitates were washed with 70% ethanol and collected. The precipitates were again dissolved in water and purity of RNA was determined by measuring OD ratio 260/280 (>1.8) and 230/260 (<0.45). The total RNA thus obtained was then purified by using the mRNA isolation kit for total RNA MACS™ (Miltenyi Biotech, Germany) and those containing poly A+ were concentrated.

cDNA synthesis

5 to 10 μg of this polyA+ rich RNA, 5 μg of the first-strand primer containing an BamHI site 5'-$(GA)_5$AGGATCCAAGAGCTC$(T)_{16}$VN-3') (SEQ ID NO:4) and 11.2 μl 80% glycerol have been combined in a total volume of 24 μl. RNA/primer mixture was denatured at 65°C for 10 min. In parallel, we combined, in a final volume of 76 μl, 18.2 μl 5X first-strand synthesis buffer, 9.1 μl 0.1 M DTT, 6.0 μl 10 mM (each) dTTP, dGTP, dATP, and 5-methyl-dCTP (instead of dCTP), 29.6 μl saturated trehalose (approximately 80%, low metal content; Fluka Biochemika), and 10.0 μl Superscript II reverse transcriptase (200 U/μl). We placed 1.0 μl [α-^{32}P]dGTP in a third tube. The mRNA, glycerol, and primers were mixed on ice with the solution containing the Superscript, and an aliquot (20%) was quickly added to the tube containing the [α-^{32}P]dGTP. First-strand cDNA syntheses were performed in a thermocycler with a heated lid (e.g., MJ Research) according to the following program: step 1, 45°C for 2

min; step 2, gradient annealing: cool to 35°C over 1 min; step 3, complete annealing: 35°C for 2 min; step 4, 50°C for 5 min; step 5, increase to 60°C at 0.1°C per second; step 6, 55°C for 2 min; step 7, 60°C for 2 min; step 8, return to step 6 for 10 additional cycles. Incorporation of radioactivity allowed the estimation of the yield of cDNA (Carninci and Hayashizaki, 1999). The cDNA was treated with proteinase K, phenol/chloroform and chloroform-extracted, and ethanol-precipitated by using ammonium acetate as the salt (Carninci and Hayashizaki, 1999).

mRNA biotinylation

Before biotinylation, the diol group of the cap and 3' end of mRNA was oxidized in a reaction solution in a final volume of 50 μ l, containing the resuspended mRNA/cDNA comprising first-strand cDNA, 66 mM sodium acetate (pH 4.5), and 5 mM $NaIO_4$. Samples were incubated on ice in the dark for 45 min. mRNA/cDNA hybrids were then precipitated by adding 0.5 μ1 of 10% SDS, 11 μ1 NaCl, and 61 μ1 of isopropanol. After incubation in the dark on ice for 45 min, the sample was centrifuged for 10 min at 15,000 rpm. Finally the mRNA/cDNA hybrids were rinsed twice with 70% ethanol and resuspended in 50 μ1 of water. Subsequently, the cap was biotinylated in a final volume reaction solution of 210 μ1 by adding 5 μ1 1 M sodium acetate (pH 6.1), 5 μ1 10% SDS, and 150 μ1 of 10 mM biotin hydrazide long-arm (Vector Biosystem). After overnight (13 hours) incubation at room temperature, the mRNA/cDNA hybrids were precipitated by adding 75 μ1 1 M sodium acetate (pH 6.1), 5 μ1 5 M NaCl, and 750 μ1 absolute ethanol and incubated on ice for 1 hour. The mRNA/cDNA hybrids were pelleted by centrifugation at 15,000 rpm for 10 min; then the pellet was washed once with

70% ethanol and once with 80% ethanol. The mRNA/cDNA hybrids were then resuspended in 70 μl 0.1X TE (1 mM Tris [pH 7.5], 0.1 mM EDTA).

Adsorption and release of full-length cDNA

500 μl of MPG-streptavidin beads and 100 μg DNA-free tRNA were combined and the obtained mixture incubated on ice for 30 min with occasional mixing. The beads were separated by using a magnetic stand for 3 minutes, and the supernatant was removed. The beads were then washed three times with 500 μL washing/binding solution (2 M NaCl, 50 mM EDTA [pH 8.0]).

At the same time, 1 unit of RNase I (Promega) per 1 μg of starting material mRNA was added to the mRNA/cDNA hybrid sample in the buffer attached to the product (final volume, 200 μl); the sample was incubated at 37°C for 15 min. To stop the reaction, the sample was put on ice and 100 μg tRNA and 100 μl of 5 M NaCl were added. To adsorb the full-coding/full-length mRNA/cDNA hybrids, the biotinylated, RNase I-treated mRNA/cDNA and the washed beads, which were resuspended in 40 μl of the washing/binding solution were combined. After mixing, the tube was gently rotated for 30 min at room temperature. Full-coding/full-length cDNA was adsorbed on the beads, and the shortened cDNAs did not. The beads were separated from the supernatant with a magnetic stand. The beads were gently washed to remove the nonspecifically adsorbed cDNAs. Two washes with washing/binding solution were performed: one with 0.4% SDS, 50 μg/ml tRNA; one with 10 mM Tris-HCl (pH 7.5), 0.2 mM EDTA, 40 μg/ml tRNA, 10 mM NaCl, and 20% glycerol; as well as with 50 μg/ml

tRNA in water.

The cDNA was released from the beads by adding 50 μl 50 mM NaOH, 5 mM EDTA and incubating for 10 min at room temperature with occasional mixing. The beads then were removed magnetically, and the eluted cDNA was transferred on ice to a tube containing 50 μl 1 M Tris-HCl, pH 7.0. The elution cycle was repeated once or twice with 50 μl-aliquots of 50 mM NaOH, 5 mM EDTA until most of the cDNA (80 to 90%, as measured by monitoring the radioactivity with a hand-held monitor) were recovered from the beads.

To remove traces of RNA, 1 μl RNase 1 (10U/μl) to the recovered cDNA on ice was quickly added; the sample was then incubated at 37°C for 10 min. The cDNA was treated with proteinase K, and then phenol/chloroform-extracted, and back-extracted. Then, the samples were concentrated by using one round of ultrafiltration with a Microcon 100 (Millipore) for 40-60 min at 2000 rpm.

CL-4B spin-column fractionation of cDNA

The cDNA samples were then treated with CL-4B chromatography (Carninci and Hayashizaki, 1999) according the manual (S-400 spin column, for example of Amersham-Pharmacia, can also be used).

cDNAs-linker ligation

Cap-Trapper full-length single strand cDNAs, prepared as above, were divided

in three different tubes. One for G-tailing, the second one for GN_5 linker and the last one for N_6/GN_5 mixed linker. An aliquot of 200 ng of cDNA were tailed with dG homopolymer as described in the prior art and used for the control cDNA library preparation (Carninci et al., Genomics, 1996).

300 ng of Cap-Trapper full-length first strand cDNAs were used as substrate for the linker-ligation using the linkers prepared as above by Gibco-BRL/Life Technologies, and cDNA libraries were constructed (shown in Figures 1 and 2).

The 300 ng of the single strand cDNA were added to 800 ng of a mixture of N_6/GN_5 linkers at proportion 1:4, and to 800 ng of GN_5 linker.

Ligation substrates (the cDNA/linker prepared as above), Solution I and Solution II (Ligation Kit, Takara) were mixed in ratio 1:2:1 and all the process were performed as described in the manual attached to the product. The reaction volume therefore was of 30 μl and contained 7.5 μl of sample, 15 μl of Solution I and 7.5 μl of Solution II. The reaction run overnight at 10℃ (Fig. 2E).

Isolation from excess linkers.

After annealing and ligation between cDNA and linker, gel filtration was carried out. 30 μl of linker-ligation samples, as above, were treated with 0.2 mg/ml proteinase K in the presence of 10 mM EDTA and 0.2% SDS. They were incubated at 45℃ for 15 min, followed by the phenol/chloroform extraction. The samples were back extracted

with 60 μl of column buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1 M NaCl, 0.1% SDS, pH 7.5). Subsequently, the samples were subjected to the spun column for gel filtration Sephacryl S 300 (Amersham Pharmacia Biotech). In the step for the spun column, the centrifugation was carried out at 400x g for 2 min. The eluted fraction was recovered and precipitated with isopropanol.

After purification step, the second strand cDNA synthesis was carried out (Fig. 2F).

To synthesize the second-strand cDNA, all purified linker-ligated samples were used. 6 μl of 10 x ExTaq buffer Takara, and 6 μl of 10 mM dNTPs and 0.5 μl [α-^{32}P]dGTP were added to the tubes in 60 μl. The samples were pre-incubated at 72℃ for 15 sec and then 0.5 μl of ExTaq were added. Then they were incubated at 72℃ for 30 min.

The samples were analyzed by alkali gel electrophoresis, that is, 0,5 μl of the samples comprising the synthesized second-strand were added to 1 μl of 6 x Alkaline dye (Sambrook, Molecular Cloning, 6.7, 6.12) in final volume of 6 μl and the electrophoresis was performed.

The electrophoresis is performed using an agarose gel containing 0.8 % of agarose, 50 mM NaOH, 5 mM EDTA and electrophoresis buffer containing 50 mM NAOH and 5 mM EDTA (Sambrook, Molecular Cloning).

The samples were purified with phenol/chloroform followed by ethanol precipitation under standard condition (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) "Molecular Cloning: A Laboratory Mannual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.).

Subsequently, cDNA was cleaved with *Bam* HI (25 U/μg of cDNAs) and *Xho* I (25 U/μg of cDNAs) at 37℃ for 1h and extracted with phenol/chloroform. The aqueous phase was purified with CL4B gel filtration spin column (Amersham Pharmacia Biotech) followed by ethanol precipitation as described in presence of 2 μg of glycogen (Carninci and Hayashizaki, 1999).

Construction of pBS IV vector

10 ng pBS SK+ (stratagene), 20μl 10x NEB buffer 2 (New England Biolabs, Inc), 20μl 10 mg/ml bovine serum albumin (NEB), 30 units Not I (NEB), 30 units Kpn I(NEB) and 10 units Xho I(NEB) were mixed in the volume of 200μl and incubated at 37 °C for 2 hours. Then, this mixture was subjected to electrophoresis at 50V for 1 hour on 0.8% SeaPlaque. agarose gel (FMC Bioproducts)/1x TAE buffer/0.5μg/ml ethidium bromide (8cm x 8cm) in 1x TAE/0.5μg/ml ethidium bromide buffer to separate a long plasmid part from a short DNA section (Molecular Cloning). The long plasmid part was cut out from the gel and the gel was transferred to a tube. Cleaved plasmid was extracted and purified by the GENECLEAN II™ Kit (Bio 101 Inc.). Concentration and purity of the plasmid were checked by agarose gel electrophoresis, comparing with standard plasmid, which concentration was already known.

Double strand oligonucleotide preparation

The oligonucleotides used were custom-synthesized (Life Technologies-Life Tech Oriental, Tokyo, Japan) and then purified by using denaturing polyacrylamide gel electrophoresis in order to remove contaminants (Maniatis etc.). The Not/Kpn double strand oligonucleotide was prepared by mixing the following two single strand oligonucleotides:

one strand (Upper-strand)

(5'GGCCGCATAACTTCGTATAGCATACATTATACGAAGTTATGGATCAGGCCAAA TCGGCCGAGCTCGAATTCGTCGACGAGAGACTGCAGGAGAGAGGATCCGGTA C-3')(SEQ ID NO:6); and

the other strand (Lower-strand)

(5'CGGATCCTCTCTCCTGCAGTCTCTCGTCGACGAATTCGAGCTCGGCCGATTT GGCCTGATCCATAACTTCGTATAATGTATGCTATACGAAGTTATGC -3')(SEQ ID NO:7) in NaCl (final concentration, 100 mM).

This mixture was then incubated at 65 ° C for 5 min, 45 ° C for 5 min, 37 ° C for 10 min, and 25° C for 10 min.

Vector-Oligonucleotide ligation

100 ng of this plasmid, which has Kpn I and Not I sites at the end were mixed with 3 ng of a Not/ Kpn double strand oligonucleotide, 1 μl 10 x ligation buffer (NEB) and T4 DNA ligase (NEB) in 10 μl.

Cell transformation

A tube comprising the ligation sample was then incubated overnight at 16 °C. The ligation sample was mixed with 250 mM NaCl, 1 μg glycogen and isopropanol, then precipitated to remove buffer. Then, it was dissolved with 10 μl sterile water. 1 μl of the obtained sample was supplied to transform to a suspension of *E.coli* cells DH10B (Life tech oriental) by electroporation (following the Protocol of the Manufacturer). The transformed cell was selected on LB plate containing 100 μg/ml ampicillin. The ampicillin-resistant clone was cultured in the LB liquid medium containing 100 μg/ml ampicillin at 37 °C for 16 h with shaking. The recombinant plasmid used for the insertion of the synthesized double strand oligonucleotide was purified by alkali-SDS method (Molecular Cloning). The inserted sequence was confirmed with the M13 forward primer (SEQ ID NO:5) and the Big dye kit by ABI377 DNA sequencer (PE-Applied BioSystems).

Vector preparation

10 μg of the modified pBS SK (+) plasmid obtained as above (named pBS IV) were mixed with 20 μl 10x Bam HI buffer, 20 μl 10 mg/ml bovine serum albumin (New England Biolabs, Inc), 30 units BamH I (New England Biolabs, Inc), 30 units Sal I(New England Biolabs, Inc), adjusted to a volume of 200 μl and incubated at 37 °C for 1.5 hours. Then, 10 units Pst I (New England Biolabs, Inc) were added to the mixture in a tube and incubated for 30 min. Furthermore, dephosphorylation of the plasmid end was carried out by 0.5 units thermo-sensitive alkaline phosphatase TsAP (Life Technologies).

TsAP makes background depending on partial cut plasmid low. Dephosphorylated ends cannot ligate each other. The tube was incubated at 37 °C for 30 min. To inactivate TsAP, EDTA (final concentration of 20 mM) was added and incubated at 65 °C for 30 min. The restrict enzyme/TsAP treated plasmid was separated as described above in the vector construction step. The band corresponding to linear plasmid, which has Bam HI and Sal I site at the end, was cut out from the gel and sliced to small pieces. It was put in a tube containing 500 μl 1x β agarase buffer (NEB) and left on ice for 30 min. The buffer was changed once and left on ice more 30 min.

This tube was incubated at 65 °C for 10 min to melt the gel. β-agarase buffer was added to bring the solution to 100 μl. Then, it was cooled at 45 °C for 3 min and β-agarase (NEB) was added to the concentration of 3U/100 μl reaction. This reaction solution was incubated at 45 °C for 6 h. 10 μl of 5M NaCl and 100 μl of phenol/chloroform were added to the tube. The tube was inverted gently for 5 min and centrifuged at 15 krpm for 3 min at room temperature. The aqueous phase was recovered and followed by chloroform extraction and isopropanol precipitation. The tube was centrifuged at 15 krpm for 10 min at 4 °C and the obtained pellet was washed with 80% ethanol twice. Finally, the pellet was dissolved with sterile water to a final concentration of 100 ng/μl. The vector concentration and purity were checked by agarose gel electrophoresis, comparing with standard plasmid, which concentration was already known.

Cloning

10 ng of cDNA, obtained in the previous step, were ligated overnight to 190 ng of the aforementioned modified vector pBluescript KS (+)(Stratagene).

The cDNA-vector ligations were precipitated with EtOH 2.5 times volume. The samples were introduced into *E.coli* DH10B (Gibco BRL) by electroporation. The transformed cells were applied on LB plate containing 100 μg/ml ampicillin and cultured overnight at 37℃. 36 colonies were picked up randomly and cultured in LB ampicillin (100 μg/ml) liquid medium overnight at 37℃. Recombinant plasmids were extracted from 3 cultures (Sambrook et al., 1989). These 3 purified plasmids were sequenced from 5' end with the M13 forward primer TGTAAAACGACGGCCAGT (SEQ ID NO:5) with the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE – ABI) by using the ABI3700 DNA sequencer (PE-Applied BioSystems) according to the Kit Manual Instruction.

Example 3

Efficiency of ligation

In the ligation of linkers prepared as above to the mouse liver derived cDNAs, 2 μg of mixed linker (N_6: GN_5 = 1: 4) were ligated against various amounts of cDNAs (1 μg, 0.5 μg and 0.2 μg respectively in Lanes 1, 2 and 3 of Figure 4). Then, 50 ng of ligated cDNAs were used for the second strand synthesis and analyzed by alkaline gel electrophoresis. All electrophoresis patterns and incorporation-rate were the same (Lanes 1-3), suggesting that 2 μg of linker were efficiently ligated to any of the different amount of cDNAs. If the linker amount was not appropriate, excessively expressed

cDNA bands would be shifted up to twice size by forming hairpin structure. Instead, the first-strand cDNA and all second-strand cDNAs showed the same pattern (the same size).

Example 4

Efficiency of linker-ligation full-length cDNAs preparation

Liver mouse cDNA libraries were prepared in the same way as above described using the CAP-trapper technology.

Whether those prepared by the linker method as described in the above example are full-length cDNA was checked by confirming the presence of ATG starting codon after sequencing step. In fact, those cDNAs containing the full-coding sequence from the starting ATG were accepted as full-length cDNAs. The 5' sequences were compared with the public nucleotide database using BLAST (Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J., 1990, "Basic local alignment search tool." J. Mol. Biol. 215:403-410).

Nucleotide Sequences were determined using Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE – ABI) and the Perkin Elmer-Applied Biosystems ABI 3700 according to the Kit Manual Instructions.

Sequencing primer used is the M13 primer on the 5' side (SEQ ID NO:5).

The data are reported in Table 1. The presence of ATG (from which the transcription starts) has been marked with the corresponding position. For example, with reference to

sample 2.01, the adenosine of ATG codon has the position 63; this indicates that this cDNA sequence has 62 bp 5'-UTR.

TABLE 1

Sample code	Gene	Public db ID	RIKEN db ID	Existence and position of ATG
1) N6/GN5 mixed linker (1:4)				
2.01	*M. musculus* alpha fetoprotein	gi|6680667|	ZX00047P08	63
2.05	Human cDNA FLJ10604 fis	gi|7022741|	ZX00036I09	172
2.22	Mouse mMCM2	gi|2381484|	——	37
2.25	*M. musculus* epsilon 14-3-3 isoform	gi|57965|	——	104
2.26	*M. musculus* SH2-containing inositol phosphatase SHIP	gi|1255351|	——	105
2.28	*M. musculus* calmodulin 3	gi|6680833|	R000011E16	176
2.30	*M. musculus* EAT/MCL-1	gi|2522268|	——	61
2.31	*M. musculus* heterogenous nuclear ribonucleoprotein U	gi|3329495|	——	199
2.33	*H. sapiens* hDj9	gi|6567165|	ZX00052D18	166
2.36	*M. musculus* Lsc (lsc) oncogene	gi|1389755|	ZX00048L03	214
2) GN5 linker				
3.07	*M. musculus* ornithine decarboxylase antizyme	gi|1279827|	ZX00047L07	81
3.09	*R. norvegicus* guanosine monophosphate reductase	gi|3907578|	R000001H02	131
3.10	Mouse calmodulin	gi|469421|	R000011E16	192
3.13	*M. musculus* protein phosphatase 5	gi|2407636|	——	89
3.16	*Rattus sp.* calcium-regulated heat stable protein CRHSP-24	gi|4583308|	R000016L20	59
3.20	Murine mRNA with homology to yeast L29 ribosomal protein	gi|50320|	ZX00047I16	30
3.21	*M. musculus* ribosomal protein S3	gi|439521|	ZX00048P23	39
3.23	*M. musculus* melanome X-actin	gi|6671508|	ZX00035N19	88
3.29	*H. sapiens* CGI-47 protein	gi|4929562|	ZX00048M14	199
3.33	*M. musculus* mini chromosome maintenance deficient 6	gi|6678831|	——	131
3.34	*M. musculus* membrane protein TMS-1	gi|5853318|	R000009P22	60

These data show that cDNAs are efficiently prepared and sequenced using the linker methodology according to the present invention.

The sequencing of clone 2.05 of Table 1 is shown in the chart of Figure 6.

The sequencing of clone 3.07 of Table 1 is shown in the chart of Figure 7.

Advantages of the CAP-Trapping-linker versus the conventional G-tailing CAP-Trapping

I) Control of G-tail length has been difficult over years. To anneal the second strand primer, cDNA clones have at least 11 of dGs with an average of 13-15 dGs. Although G-tailing reaction is self-limiting to 15-30 nt (Hyone-Myong Eun, 1996, Enzymology Primer for Recombinant DNA Technology, page 477), G-stretches longer than about 20 bases, which were often obtained, caused a dramatic decrease of sequencing yield or in the worst case failure, while shorter G stretches impaired long sequence reading (see Figure 5). During sequencing, long G-stretches may interact with surrounding sequences and form very strong secondary structures. This may be problematic in case of interactions with 5' UTRs that are typically GC rich. This is especially serious in full-length cDNAs synthesis like in case of Cap-Trapping libraries.

The method using the linker according the present invention, different from conventional methods, does not have such a drawback (even if the random variable portion comprises G, there are statistically within a small number) and can allow an efficient sequencing (Figure 6 and 7).

II) G-stretch is expected to affect the efficiency of translation in case of functional studies, for instance where protein expression is required, such as in expression cloning (King RW, Lustig KD, Stukenberg PT, McGarry TJ, Kirschner MW. Expression cloning in the test tube. Science. 1997;277:973-4.). On the other hand, the linker sequence of the present invention does not inhibit transcription and translation.

INDUSTRIAL APPLICABILITY

The present invention provides a novel and efficient method for the preparation of cDNA libraries. More specifically, the present invention provides a novel linker instead of G-tailing, which can be utilized in a method for the preparation of cDNA libraries and to provide a method for the preparation of cDNA libraries using said linker.

CLAIMS

1. A linker or a population of linkers comprising an oligonucleotide fixed portion and an oligonucleotide variable portion, the variable portion being represented as Formula (N)n wherein N is A, C, G, T or U, or their derivatives, and n is an integer equal to or higher than 1, and if n is an integer equal to or higher than 2, the nucleotides (N) may be the same or different from each other.

2. The linker of claim 1, wherein the variable portion is prepared at random.

3. The linker according to claim 1, wherein n is an integer ranging from 1 to 10.

4. The linker according to claim 1, wherein n is an integer ranging from 4 to 8.

5. The linker according to claim 1, wherein n is 5 or 6.

6. The linker according to any one of claims 1 to 5, wherein the variable portion is $(G)_m(N)_{n-m}$, and m is an integer ranging from 1 to 3.

7. The linker according to any one of claims 1 to 6, wherein the variable portion is GN_4, GN_5, G_2N_3, G_2N_4, G_3N_2, G_3N_3, N_5 or N_6.

8. The linker according to any one of claims 1 to 7, wherein the oligonucleotide fixed portion is a single strand or double strand oligonucleotide.

9. The linker according to claim 8, wherein the fixed portion is a double strand oligonucleotide and the single strand variable portion (N)n protrudes outside the double strand fixed portion.

10. The linker according to any one of claims 1 to 9, which is a single strand oligonucleotide having the sequence of SEQ ID NO:1 or SEQ ID NO:3.

11. The linker according to any one of claims 1 to 9, which is a double strand

oligonucleotide having the sequence SEQ ID NO:1/SEQ ID NO:2 or SEQ ID NO:3/SEQ ID NO:2.

12. A double strand oligonucleotide linker or a population of linkers, comprising

a) a first oligonucleotide single strand comprising an oligonucleotide single strand fixed portion and an oligonucleotide single strand variable portion, and

b) a second oligonucleotide single strand comprising an oligonucleotide single strand fixed portion annealed to the complementary first oligonucleotide single strand fixed portion (a),

so that the variable portion protrudes outside the double strand fixed portion of the linker; and wherein the variable portion is represented by the formula (N)n wherein N is A, C, G, T or U, or their derivatives and n is an integer equal to or higher than 1.

13. The linker of claim 12, wherein the variable portion is prepared at random.

14. The linker of claim 12, wherein n is an integer ranging from 1 to 10.

15. The linker of claim 12, wherein n is an integer ranging from 4 to 8.

16. The linker of claim 12, wherein n is 5 or 6.

17. The linker according to any one of claims 12 to 16, wherein the variable portion is (G)m(N)n-m, and m is from 1 to 3.

18. The linker according to any one of claims 12 to 17, wherein the variable portion is GN_4, GN_5, G_2N_3, G_2N_4, G_3N_2, G_3N_3, N_5 or N_6.

19. The linker according to any one of claims 12 to 18, which has the sequence SEQ ID NO:1/SEQ ID NO:2 or SEQ ID NO:3/SEQ ID NO:2.

20. A population of linkers comprising at least two linkers according to any one of claims 1 to 19.

21. A population of linkers comprising at least two subpopulations of linkers according to any one of claims 1 to 20.

22. The population of linkers according to any one of claims 1 to 21, wherein the fixed portions comprised in all linkers are an oligonucleotide portion having the same sequence.

23. The population of linkers according to any one of claims 1 to 22, comprising two or more subpopulations of linkers, wherein one subpopulation of linkers comprise linkers in which the fixed portion is an oligonucleotide portion having the same sequence, and other subpopulations of linkers which differ each other in the fixed portion sequence.

24. The population of linkers according to any one of claims1 to 23, wherein the variable portion of the linkers is synthesized at random.

25. The population of linkers according to any one of claims 1 to 24, wherein the sequence of variable portion comprised in each linker differs from each other.

26. The population of linkers according to any one of claims 1 to 25, which is a mixture of the (N)n linker and the (G)m(N)n-m linker, represented as (N)n/(G)m(N)n-m mixture.

27. The population of linkers according to any one of claims 1 to 26, which is a mixture of two or more kinds of linkers having any variable portion selected from the group consisting of GN_4, GN_5, G_2N_3, G_2N_4, G_3N_2, G_3N_3, N_5 or N_6.

28. The population of linkers according to claim 27, which is a N_6/GN_5 mixture, N_6/G_2N_4 mixture or N_6/G_3N_3 mixture.

29. The population of linkers according to claim 28, wherein the mixing ratio ranges from 0:1 to 1:0.

30. The population of linkers according to claim 29, wherein the mixing ratio is 1:4.

31. The linker or population of linkers according to any one of claims 1 to 30, which is prepared by;

(a) synthesizing a first oligonucleotide single strand comprising an oligonucleotide single strand fixed portion and an oligonucleotide single strand variable portion,

(b) synthesizing a second oligonucleotide single strand comprising an oligonucleotide single strand fixed portion complementary to the first oligonucleotide single strand fixed portion (a), and

(c) annealing the first oligonucleotide strand (a) to the second oligonucleotide strand (b) so that the variable portion protrudes outside the double strand fixed linker portion.

32. The linker or population of linkers according to any one of claims 1 to 31, wherein the fixed portion has one or more sequences of restriction enzyme sites, recombinational sites, RNA polymerase promoter sites, markers or tags.

33. The linker or population of linkers according to claim 32, wherein the restriction enzyme site is BamHI, XhoI, SstI, SalI or NotI.

34. The linker or population of linkers according to claim 32, wherein the recombinational site is attB, cre-lox or Gateway™.

35. The linker or population of linkers according to claim 32, wherein the polymerase promoter site is selected from the group consisting of T7, T3, K11, SP6 and BA14 RNA polymerase.

36. The linker or population of linkers according to claim 32, wherein the markers

are specific sequences or plural oligonucleotides

37. The linker or population of linkers according to claim 32, wherein the tag binds to one or both ends of two strands of the fixed portion, opposite to the variable portion, and said end is a protecting group preventing the formation of binding loop between said linker and an annealed target single strand polynucleotide.

38. The linker or population of linkers according to claim 37, wherein the protecting group is a NH_2 group.

39. The linker or population of linkers according to claim 38, wherein the ends of two strands of the fixed portion of said linker, opposite to the variable portion, are bound together to form a loop so as to prevent said linker from binding to the annealed target single strand polynucleotide.

40. The population of linkers according to any one of claims 1 to 39, which is a population of linkers comprising at least one linker in which the variable oligonucleotide sequence is a specific oligonucleotide sequence, wherein the specific oligonucleotide sequence being able to specifically annealing to the end of at least one defined target polynucleotide to be selected from a population of target polynucleotides.

41. The linker or population of linkers according to any one of claims 1 to 40, which is DNA or RNA or a mixture thereof.

42. A linker-polynucleotide or population of linker-polynucleotides comprising the linker or population of linkers according to any one of claims 1 to 41 and the target first strand polynucleotide bound to the linker.

43. The linker-polynucleotide or population of linker-polynucleotides according to claim 42, wherein one end of the target first strand polynucleotide is annealed and/or

ligated to the linker.

44. The linker-polynucleotide or population of linker-polynucleotides according to claim 42, wherein one end of the target first strand polynucleotide is annealed to the variable portion of the first strand of the linker and ligated to the second strand fixed portion of the linker.

45. The linker-polynucleotide or population of linker-polynucleotides according to any one of claims 42-44, wherein the linker is ligated to the target first strand polynucleotide with ligase.

46. The linker-polynucleotide or population of linker-polynucleotides according to claim 45, wherein the ligase is a DNA ligase or RNA ligase.

47. The linker-polynucleotide or population of linker-polynucleotides according to any one of claims 42 to 46, which further comprises a second strand polynucleotide complementary to the target first strand polynucleotide.

48. The linker-polynucleotide or population of linker-polynucleotides according to claim 47, wherein the second strand polynucleotide is synthesized from a free end of the variable portion, with the variable portion acting as a primer.

49. The single strand or double strand polynucleotide or population of polynucleotides according to any one of claims 45 to 48, which is RNA or DNA, or a mixture thereof.

50. The single strand or double strand polynucleotide or population of polynucleotides according to claim 49, wherein DNA is cDNA.

51. The single strand or double strand polynucleotide or population of polynucleotides according to claim 50, which is a long strand, full-coding and/or

full-length cDNA.

52. A vector comprising the linker-polynucleotide according to any one of claims 42-51.

53. The vector according to claim 52, which is a phage, plasmid vector or vector for sequencing.

54. A method for preparing the linker or population of linkers according to any one of claims 1 to 41, which comprises the steps of:

(a) synthesizing a first oligonucleotide single strand comprising an oligonucleotide single strand fixed portion and an oligonucleotide single strand variable portion,

(b) synthesizing a second oligonucleotide single strand comprising an oligonucleotide single strand fixed portion complementary to the first oligonucleotide single strand fixed portion (a), and

(c) annealing the first oligonucleotide strand (a) to the second oligonucleotide strand (b), so that the variable portion protrudes outside the double strand fixed portion of the linker.

55. A method of binding a target single strand polynucleotide to a linker comprising the steps of:

i) preparing the linker according to any one of claims 1 to 41; and

ii) annealing the variable portion of said linker to the target single strand polynucleotide.

56. A method of binding a target single strand polynucleotide to a linker comprising the steps of:

i) preparing the linker according to any one of claims 1 to 41; and

ii) annealing the variable portion of the first strand of the linker to the target single strand polynucleotide and ligating the fixed portion of the second strand of the linker to the target single strand polynucleotide.

57. A method of binding a target single strand polynucleotide or a population of the polynucleotides to a population of linkers comprising the steps of:

i) preparing the population of linkers according to any one of claims 1 to 41; and

ii) annealing the variable portion of the population of linkers to the target single strand polynucleotide or to the population of the polynucleotides.

58. A method of binding a target single strand polynucleotide or a population of the polynucleotides to a population of linkers comprising the steps of:

i) preparing the population of linkers according to any one of claims 1 to 41; and

ii) annealing the variable portion of the first strand of said population of linkers to the target single strand polynucleotide or the population of the polynucleotides and ligating the fixed portion of the second strand of the population of linkers to the target single strand polynucleotide or the population of the polynucleotides.

59. A method of preparing a linker-polynucleotide comprising a linker and a double strand polynucleotide, comprising the steps of:

i) annealing the variable portion of the linker according to any one of claims 1 to 41 to the target first strand polynucleotide, and

ii) synthesizing the second strand polynucleotide complementary to the target single strand polynucleotide.

60. A method of preparing a linker-polynucleotide comprising a linker and a double

strand polynucleotide, comprising the steps of:

i) annealing the variable portion of the first strand of the linker according to any one of claims 1 to 41 to a target single strand polynucleotide and ligating the target single strand polynucleotide to the fixed portion of the second strand of the linker, and

ii) synthesizing the second single strand polynucleotide complementary to said target single strand polynucleotide.

61. A method of preparing a linker-polynucleotide comprising a linker and a double strand polynucleotide, comprising the steps of:

i) annealing the variable portion of the linker or a population of the linkers according to any one of claims 1 to 41 to a target single strand polynucleotide or a population of the target single strand polynucleotides, and

ii) synthesizing the second strand polynucleotide complementary to said target single strand polynucleotide or a population thereof.

62. A method of preparing a linker-polynucleotide comprising a linker and a double strand polynucleotide, comprising the steps of:

i) annealing the variable portion of the first strand of the linker or a population of the linkers according to any one of claims 1 to 41 to a target single strand polynucleotide or a population of the target single strand polynucleotides,

ii) ligating the target single strand polynucleotide or the population of the target single strand polynucleotides to the fixed portion of the second strand of the linker or the population of the linkers, and

iii) synthesizing the second single strand polynucleotide(s) complementary to said target single strand polynucleotide(s).

63. The method according to any one of claims 54-62, wherein one end of the target first strand polynucleotide is annealed to the variable portion of the linker, and ligated to the fixed portion of the second strand of said linker.

64. The method according to any one of claims 54-63, wherein the variable portion is synthesized at random.

65. The method according to any one of claims 55-64, wherein the population of linkers comprises at least one linker in which the variable oligonucleotide portion is a specific oligonucleotide portion, this specific oligonucleotide portion being complementary to the end of at least one specific polynucleotide target to be selected from the group of target polynucleotides.

66. The method according to any one of claims 55-65, wherein ligation is performed by ligase.

67. The method according to claim 66, wherein the ligase is a DNA ligase or RNA ligase.

68. The method according to claim 67, wherein the DNA ligase is T4 DNA ligase or *E. coli* DNA ligase.

69. The method according to any one of claims 66-68, wherein the ligation is performed in the presence of a ligase-stimulating agent.

70. The method according to claim 69, wherein the ligase-stimulating agent is PEG (polyethylene glycol).

71. The method according to any one of claims 55-70, wherein the linker and the target first strand polynucleotide and/or the second strand polynucleotide complementary to the target is DNA.

72. The method according to claim 71, wherein the target first strand polynucleotide or the double strand polynucleotide is a long strand, full-coding and/or full-length cDNA.

73. The method according to claim 72, wherein the first strand cDNA is obtained from the Cap trapping at the 5' end of mRNA.

74. The method according to claim 73, wherein the Cap-trapped cDNA is further normalized or subtracted before or after the ligation to linker.

75. The method according to any one of claims 55-74, which comprises the step of increasing temperature before annealing the linker to the target first strand polynucleotide and/or after synthesizing polynucleotide second strand.

76. The method according to claim 75, wherein the temperature ranges from 25 to 100℃.

77. The method according to claim 76, wherein the temperature is 65℃.

78. The method according to any one of claims 55-77, wherein at least one end of the fixed portion of the linker, opposite to the variable portion, is tagged with a protective group.

79. The method of claim 78, wherein the protective group is NH_2.

80. The method of any one of claims 55-79, wherein the ends of the two strands of the linker fixed portion are bound together by making a loop.

81. The method according to any one of claims 55-80, which further comprises the step in which said linker-polynucleotide is cleaved at both ends in restriction enzyme sites and inserted into a vector.

82. The method according to any one of claims55-80, which further comprises the step in which said linker-polynucleotide is cleaved remaining blunt ends and inserted into

a vector.

83. A method of marking a polynucleotide library and distinguishing said library, comprising the step of providing a population of linkers, comprising a fixed portion and a variable portion, wherein the fixed portion comprises at least one marker indicating a specific or defined tissue or species, and selecting and separating said library by said specific or defined marker.

84. The method according to claim 83, wherein said linker or population of linkers is the one according to any one of claims 1 to 41.

85. The method according to claim 83 or 84, wherein the fixed portions comprised in all linkers among the population of linkers are oligonucleotide portions having the same sequence.

86. A method of binding a linker or population of linkers to mRNA, which comprises the steps of:

(a) treating mRNA with phosphatase and removing phosphate groups from uncapped mRNA,

(b) treating a product of step (a) with pyrophosphatase, which removes the CAP structure from capped mRNA, and

(c) adding an RNA ligase in the presence of the linker according to any one of claims 1 to 41.

87. A method of preparing a linker-polynucleotide, which comprises the steps of:

(a) treating mRNA with phosphatase and removing phosphate groups from uncapped mRNA,

(b) treating a product of step (a) with pyrophosphatase, which removes the CAP structure from capped mRNA,

(c) adding an RNA ligase in the presence of the linker according to any one of claims 1 to 41, and

(d) adding an oligo dT and synthesizing a polynucleotide complementary to the complete sequence of said mRNA.

88. The method according to claim 87, wherein the polynucleotide complementary to mRNA is cDNA.

89. The method according to any one of claims 86-88, wherein the phosphatase is bacterial alkaline phosphatase (BAP).

90. The method according to any one of claims 86-88, wherein the phosphatase is tobacco acid pyrophosphatase (TAP).

91. The method according to any one of claims 86-90, wherein the linker of step (c) is DNA and further comprising the steps of

(e) adding RNase H,

(f) adding DNA polymerase I, and

(g) synthesizing a cDNA strand.

92. A method of binding a linker or a population of linkers according to any one of claims 1 to 41 to a target single strand polynucleotide or population of polynucleotides comprising the step of adding RNA ligase.

93. The method according to claim 92, which further comprises the addition of an oligonucleotide primer or population of primers complementary to the variable portion of the linker or population of linkers.

94. The method according to claim 93, wherein the primer comprised in the population of primers is randomly synthesized.

95. A method of preparing a second strand polynucleotide, which comprises the step according to any one of claims 92-94 and further synthesizing the second strand polynucleotide complementary to a target polynucleotide.

96. A method of preparing a DNA/RNA hybrid, which comprises the steps of:

i) providing a full-length/coding or long poly-A mRNAs,

ii) ligating and annealing said mRNA to the linker according to any one of claims 1 to 41, the linker comprising a first restriction enzyme site,

iii) annealing an oligo dT-primer comprising a second restriction enzyme site to the mRNA,

iv) synthesizing a cDNA strand,

v) isolating the hybrid by using restriction enzymes which recognize the two specific restriction enzyme sites introduced, and

vi) cloning.

97. A method of preparing a linker-polynucleotide product comprising a linker and a single strand polynucleotide, comprising annealing the variable portion of the linker according to the present invention to the target first strand polynucleotide.

Fig.1

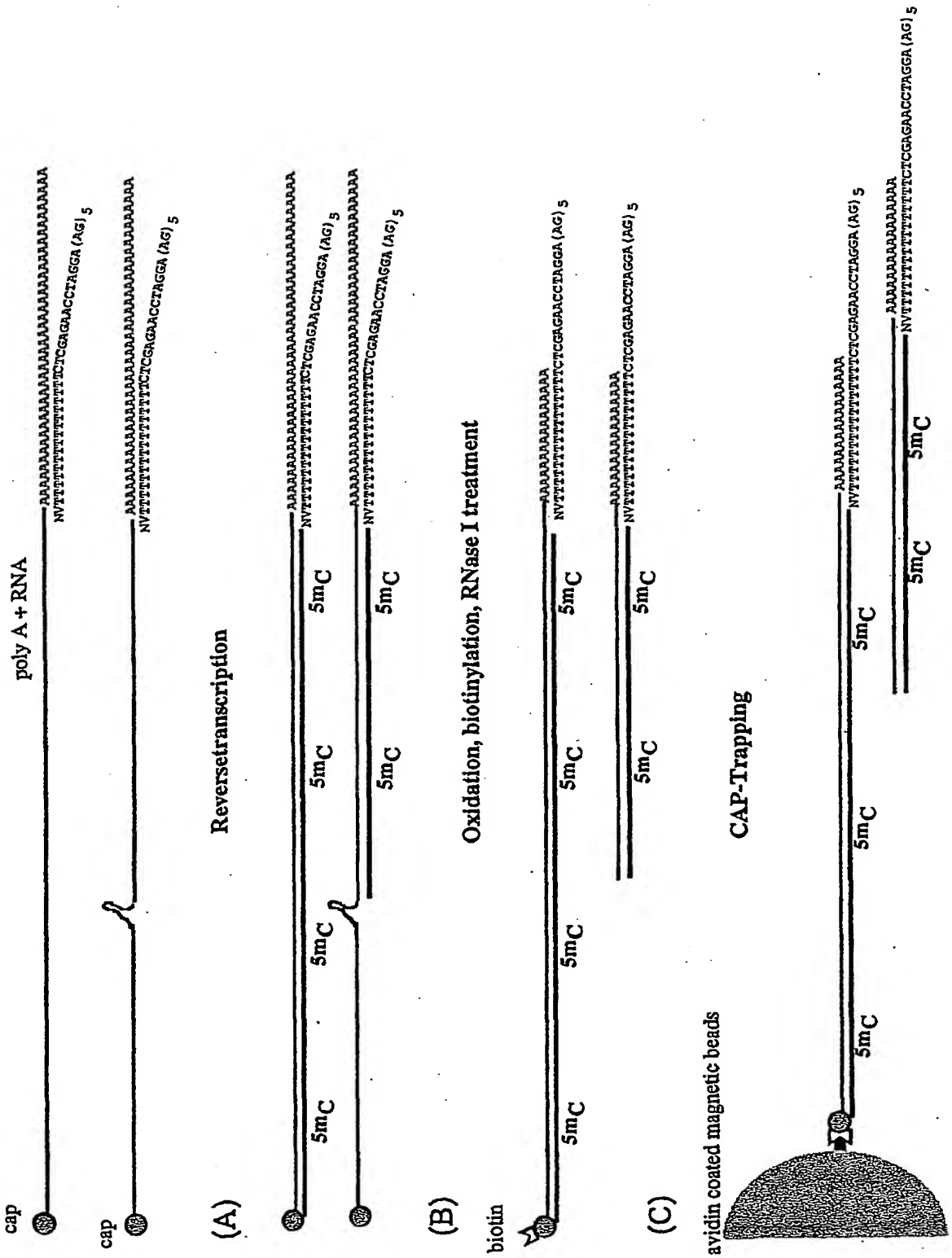

poly A + RNA
cap
cap
(A)
Reversetranscription
5mC
(B)
Oxidation, biotinylation, RNase I treatment
biotin
(C)
CAP-Trapping
avidin coated magnetic beads

Fig.2

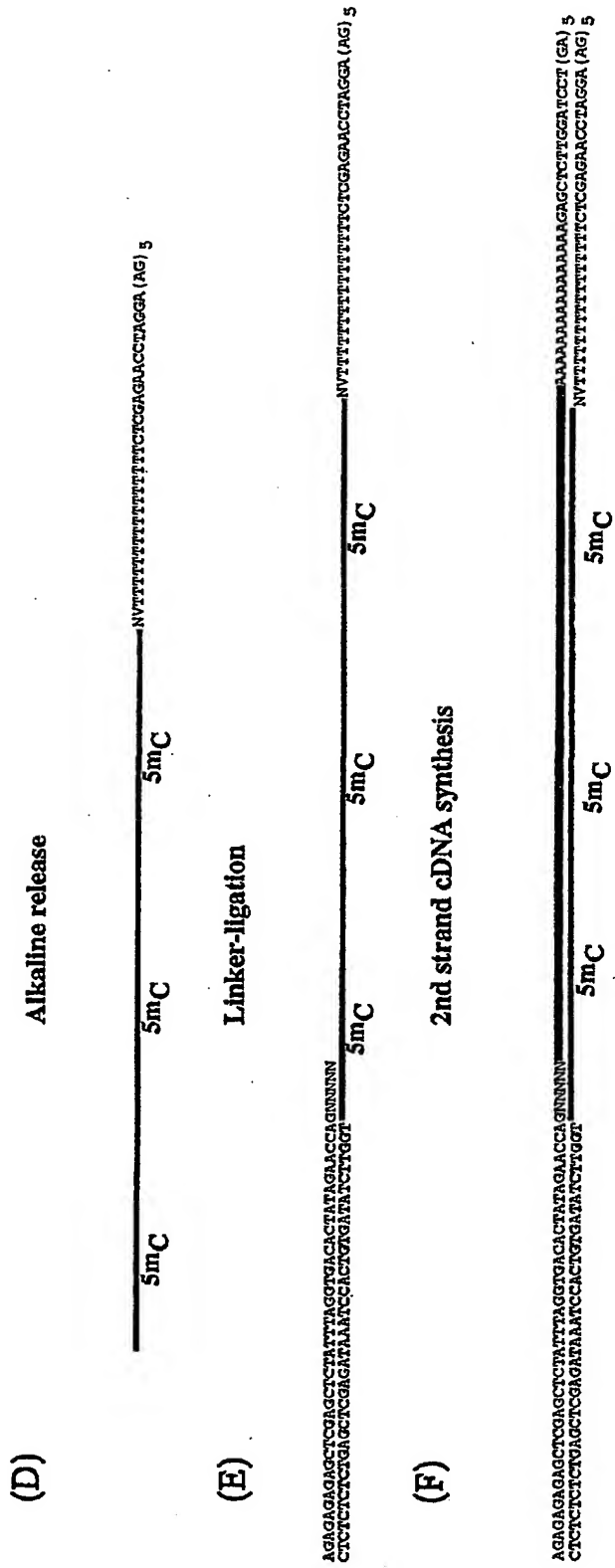

(D)
Alkaline release
(E)
Linker-ligation
(F)
2nd strand cDNA synthesis
(G)
Restriction cut
(H)
Ligation with lambda cloning vector
(I)
Packaging

Fig.3

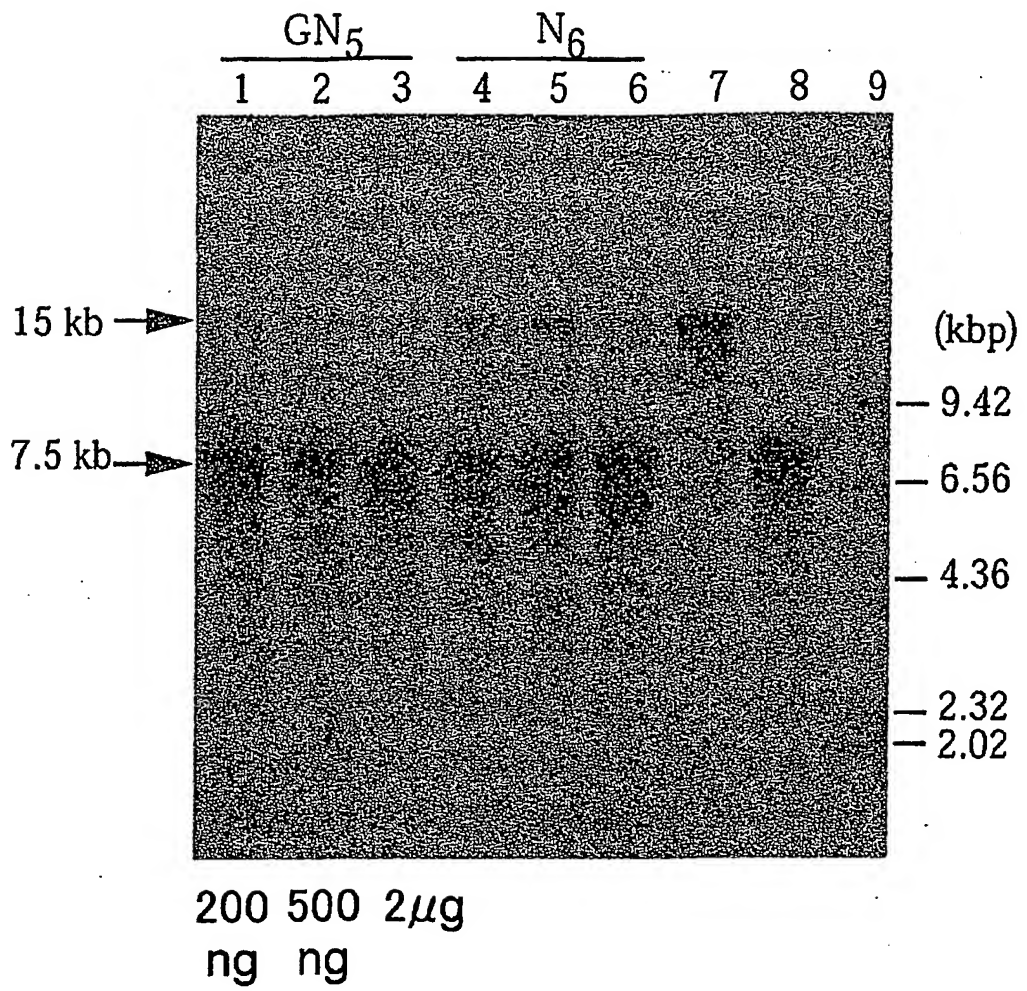

GN5
N6
1 2 3 4 5 6 7 8 9
15 kb
7.5 kb
(kbp)
9.42
6.56
4.36
2.32
2.02
200 500 2μg
ng ng

Fig.4

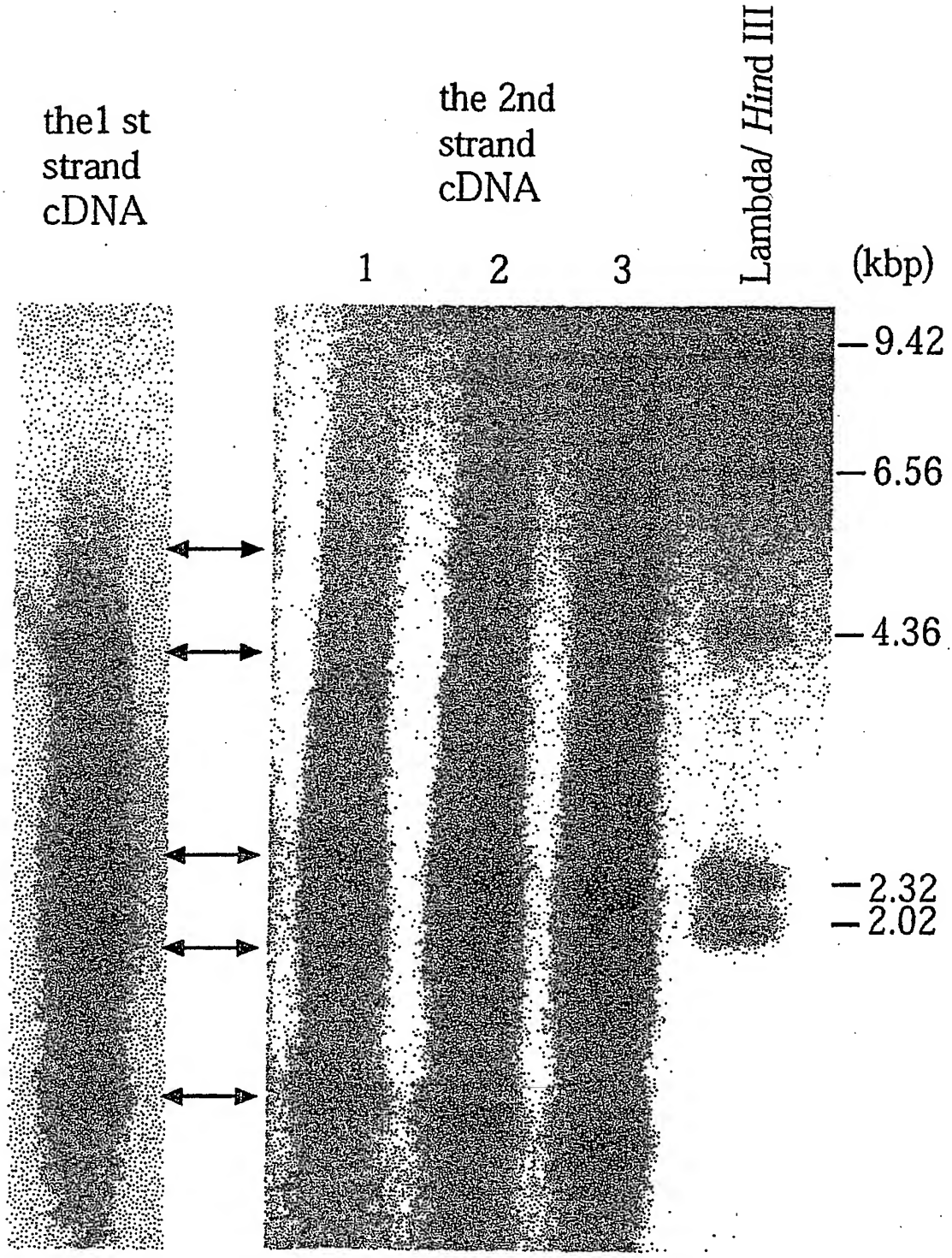

the1 st strand cDNA
the 2nd strand cDNA
Lambda/ Hind III
1
2
3
(kbp)
9.42
6.56
4.36
2.32
2.02

Fig.5(a)

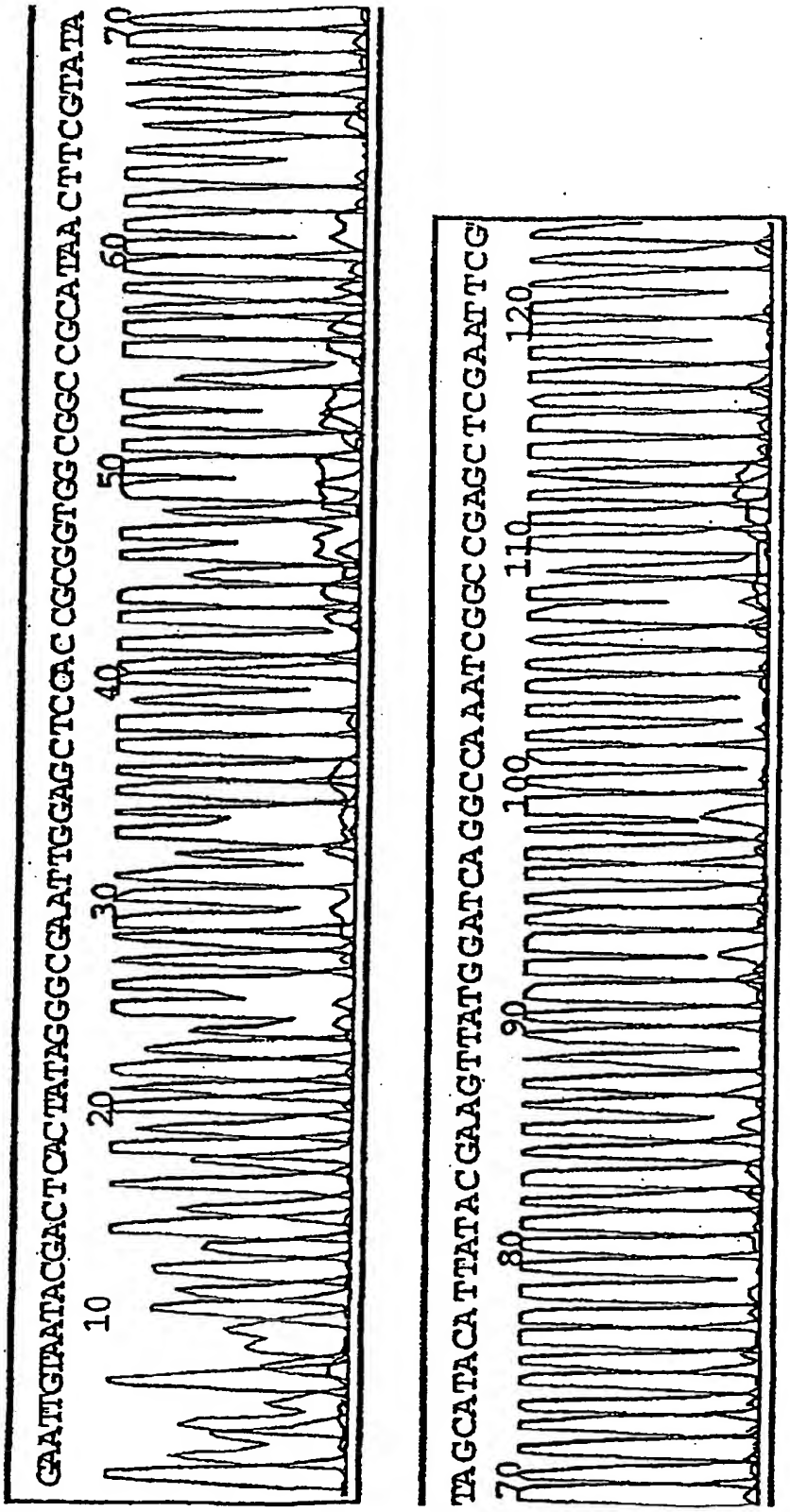

GAATTGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCATAACTTCGTATA
10
20
30
40
50
60
70
TAGCATACATTATACGAAGTTATGGATCAGGCCAAATCGGCCGAGCTCGAATTCG
70
80
90
100
110
120

Fig.5(b)

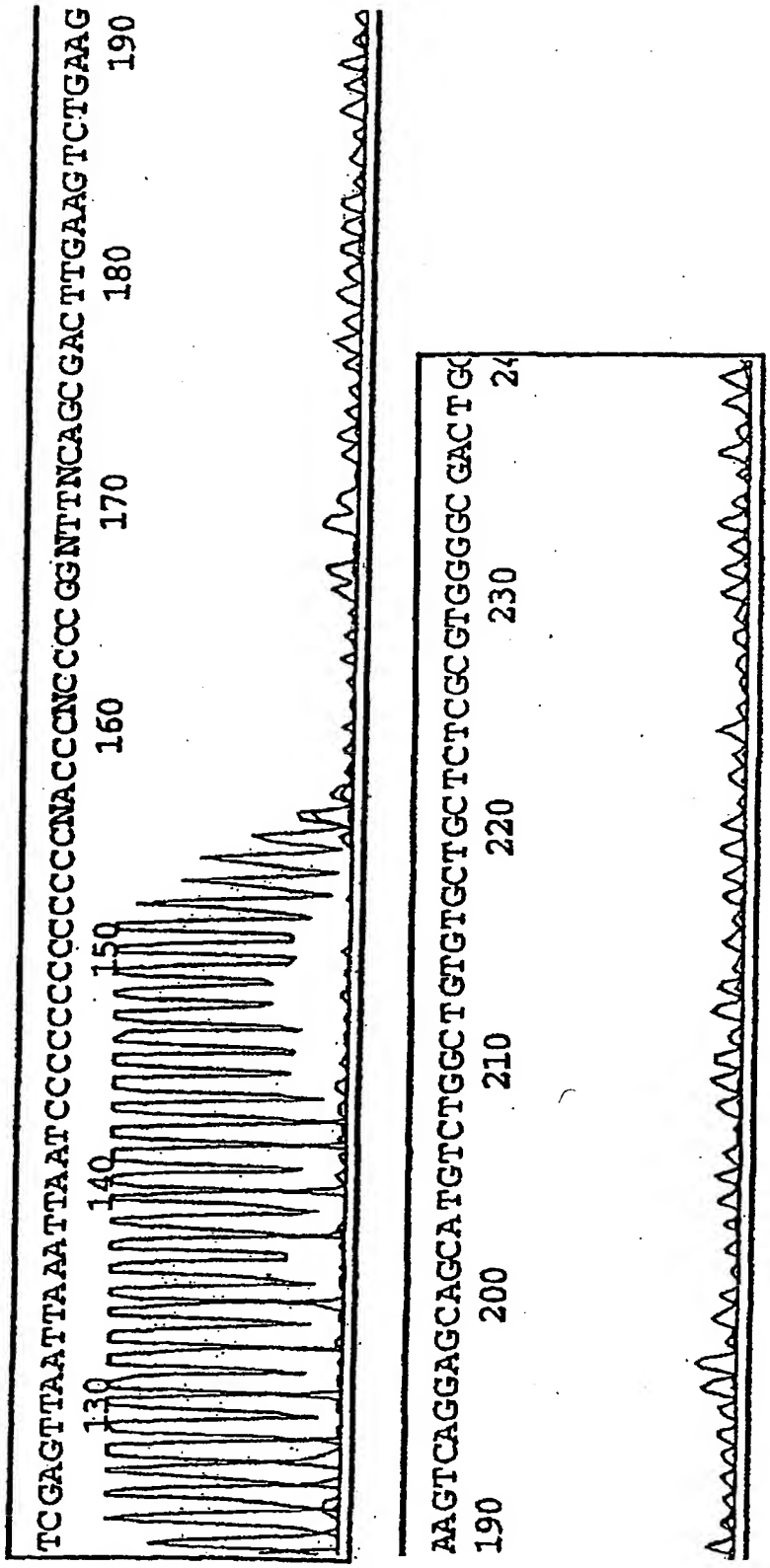

TCGAGTTAATTAATTAATCCCCCCCCCCCNACCCNCCCGGNTTNCAGCGACTTGAAGTCTGAAG
130
140
150
160
170
180
190
AAGTCAGGAGCAGCATGTCTGGCTGTGTGCTGCTCTCGCGTGGGGGCGACTG
190
200
210
220
230

Fig.5(c)

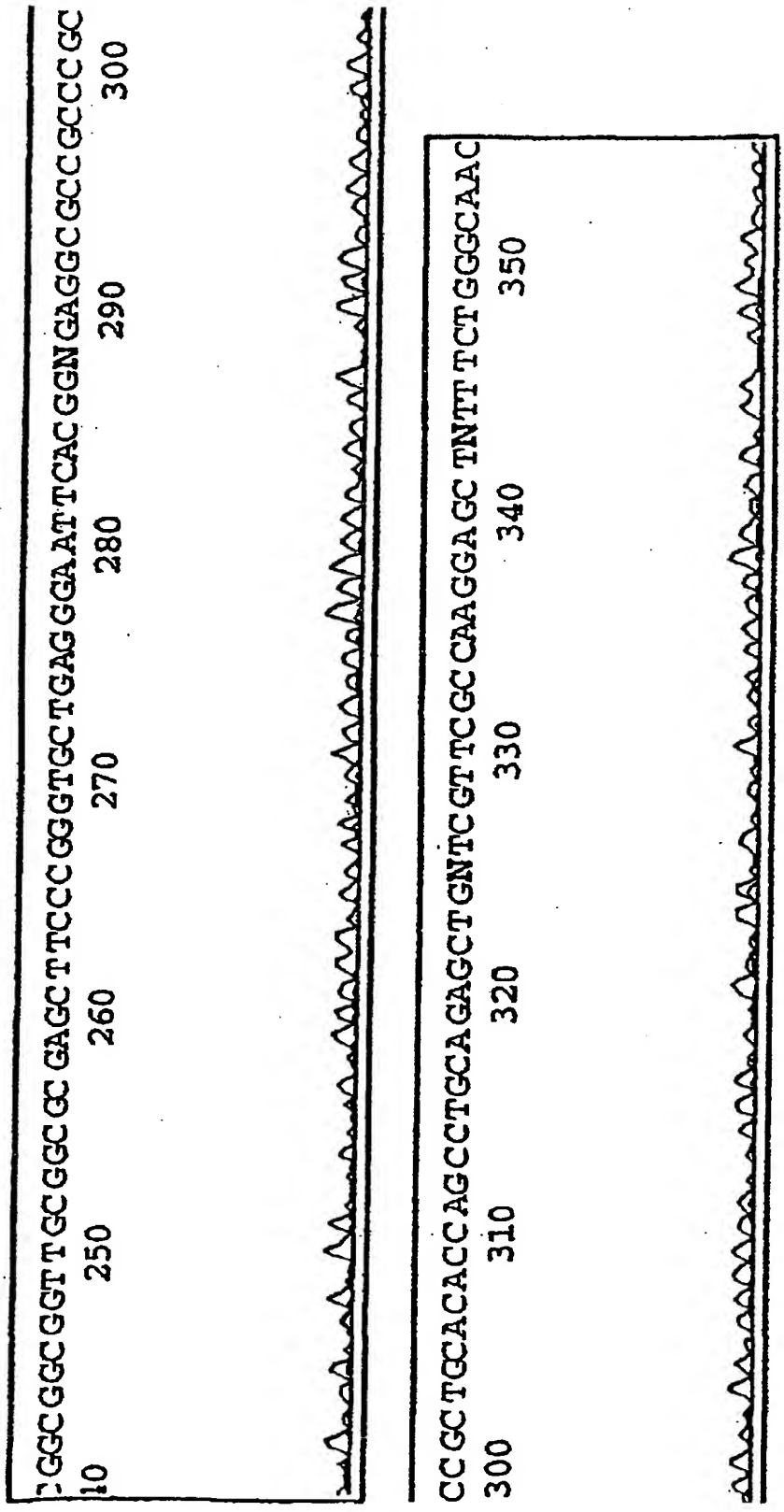

GGC GGC GGT TGC GGC GC GAG CTT CCC GGG TGC TGA GGA GGA ATT CAC GGN GAG GCC GCC GCC GC
10 250 260 270 280 290 300
CC GC TGC ACA CCA GCC TGC AGA GGC TGN TCG TTC GC CAA GGA GC TNT TCT GGG CAA C
300 310 320 330 340 350

Fig.5(d)

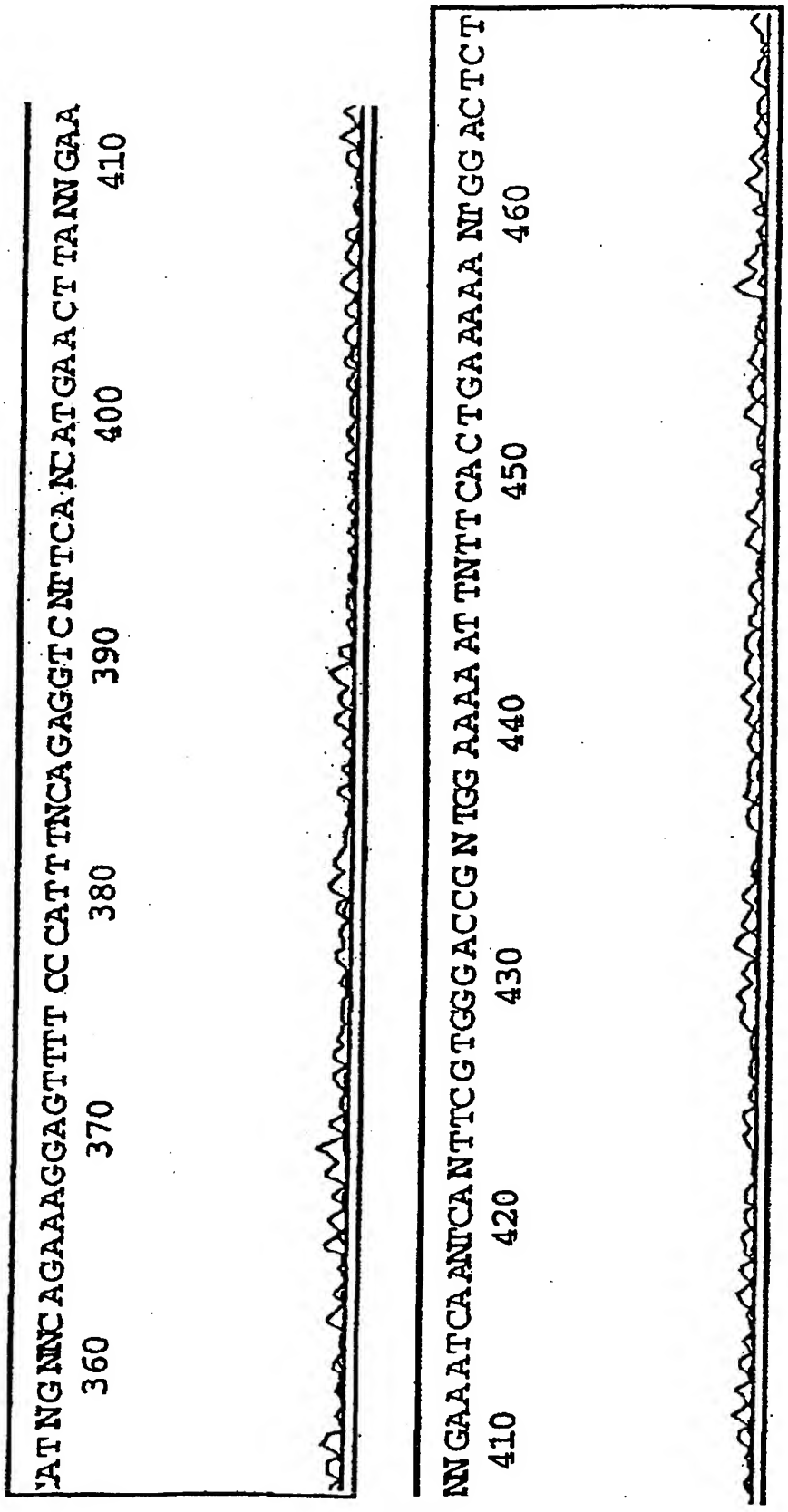

'AT NG NNC AGAAAGGAGTTTT CC CAT T TNCA GAGGT C NT TCA NC AT GA A C T TA NN GAA
360 370 380 390 400 410
NN GAA A T CA ANTCA N TTC G T GGG A CCG N TGG AAAA AT TN T T CA C T GA AAAA NT GG AC T C T
410 420 430 440 450 460

Fig.5(e)

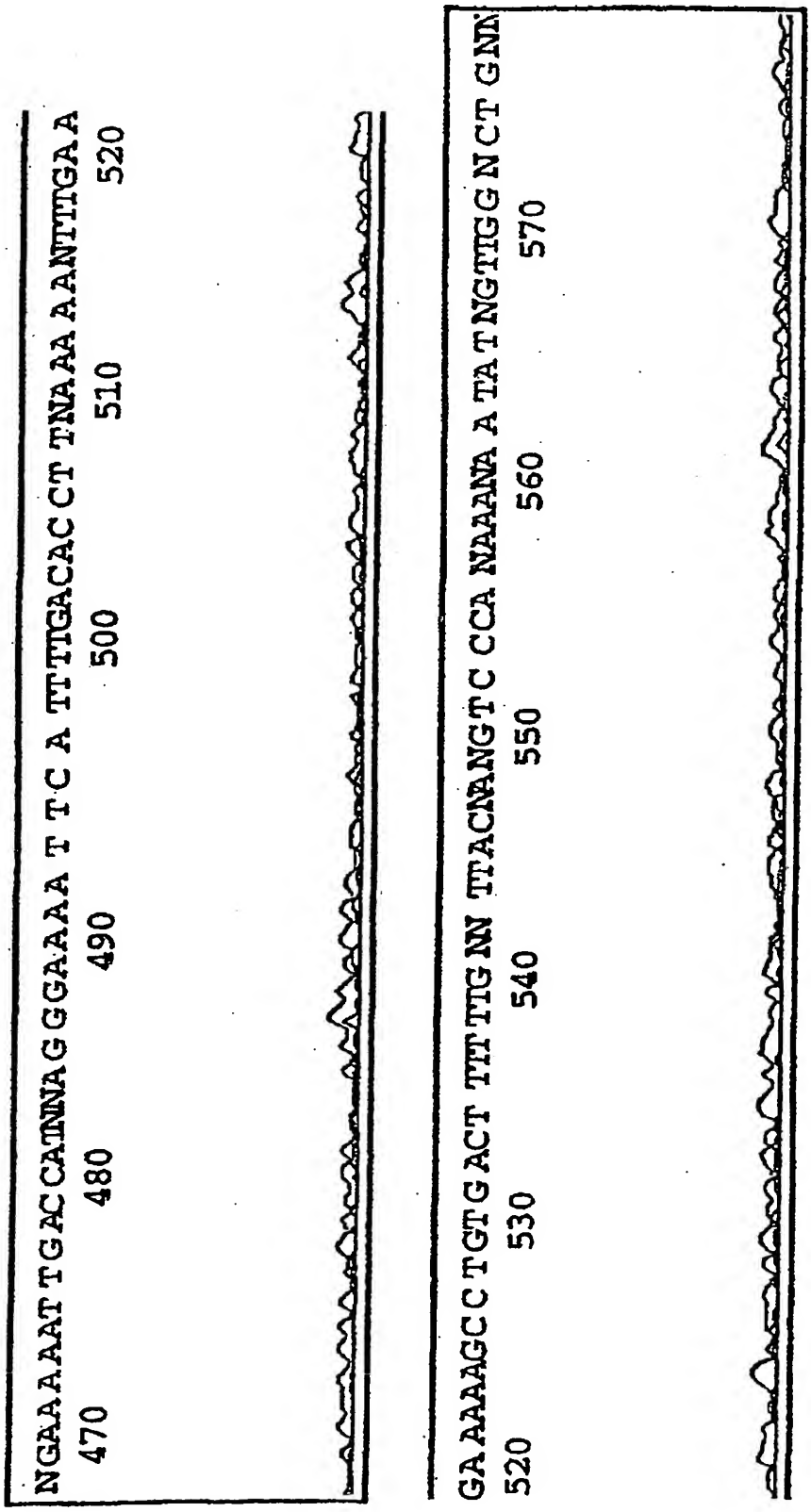

NGAAAAATTGACCATNNAGGGAAAA T TC A TTTTGACAC CT TNAAA AANTTTGAAA
470 480 490 500 510 520
GA AAAAGCC T TGTG ACT TTT TTG NN TTACNANGTC CCA NAAANA A TA T NGTTGG N CT GNN
520 530 540 550 560 570

Fig.6(a)

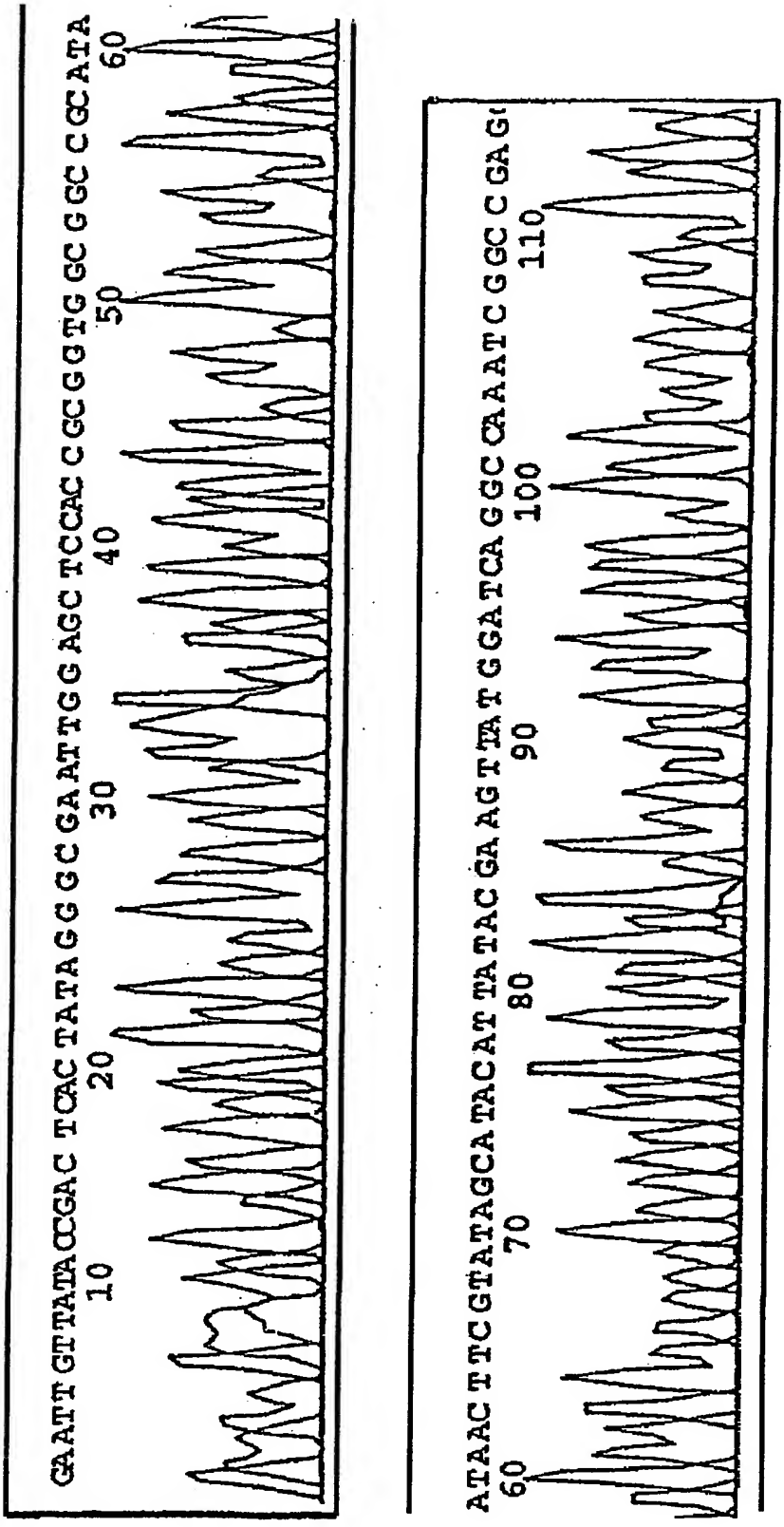

GAATTGTTATACCGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCCCATA
10
20
30
40
50
60
ATAACTTCGTATAGCATACATTATACGAAGTTATGGATCAGGCCGAAATCGGCCGAG
60
70
80
90
100
110

Fig.6(b)

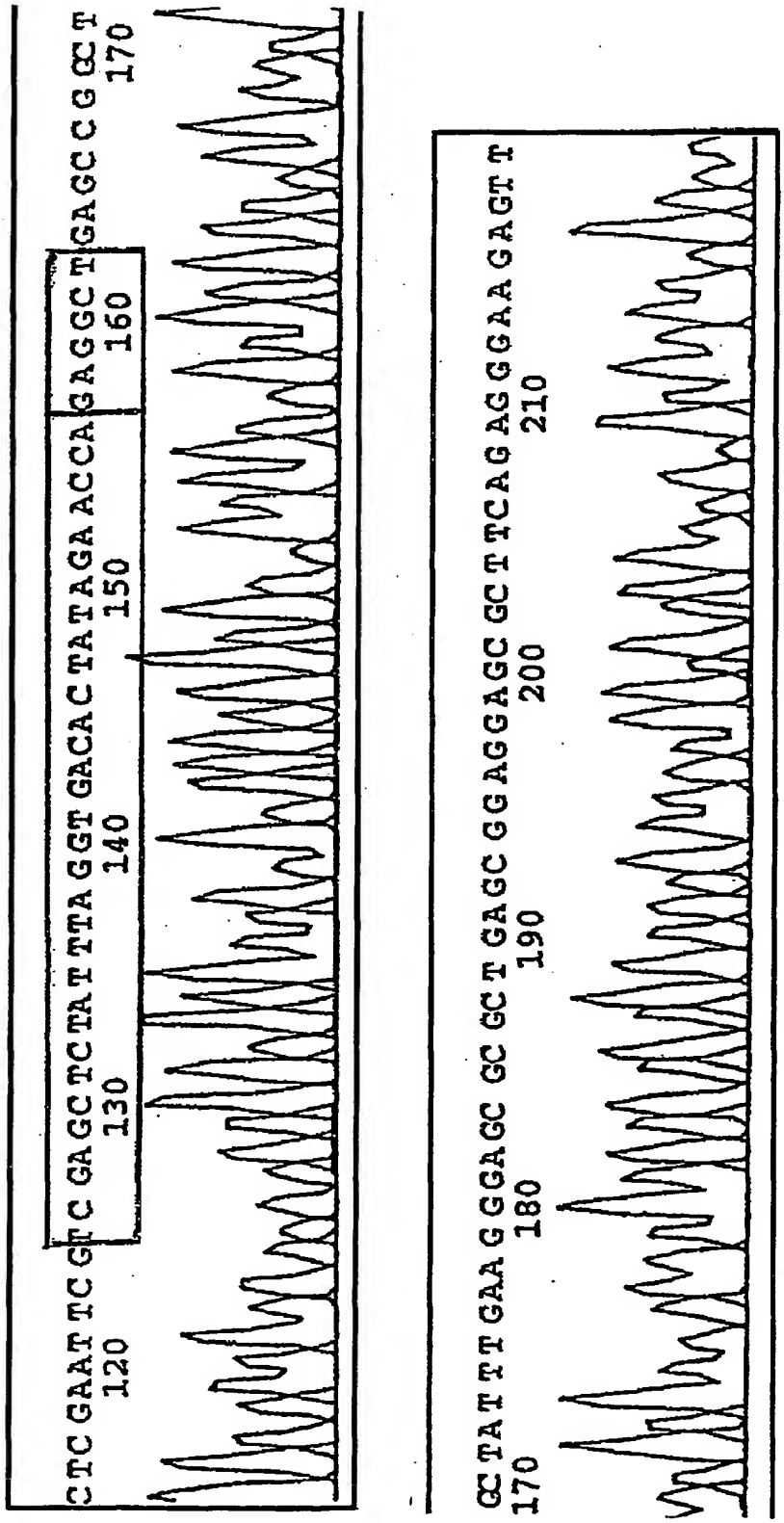

CTC GAA TTC GTC GAG CTC TAT TTA GGT GAC ACT ATA GAA CCA GAG GCT GAG CCG CT
120 130 140 150 160 170
GC TAT TT GAA G GGA GC GC T GA GC GG AG GAG C GC T TCA G AG GAA GAG T T
170 180 190 200 210

Fig.6(c)

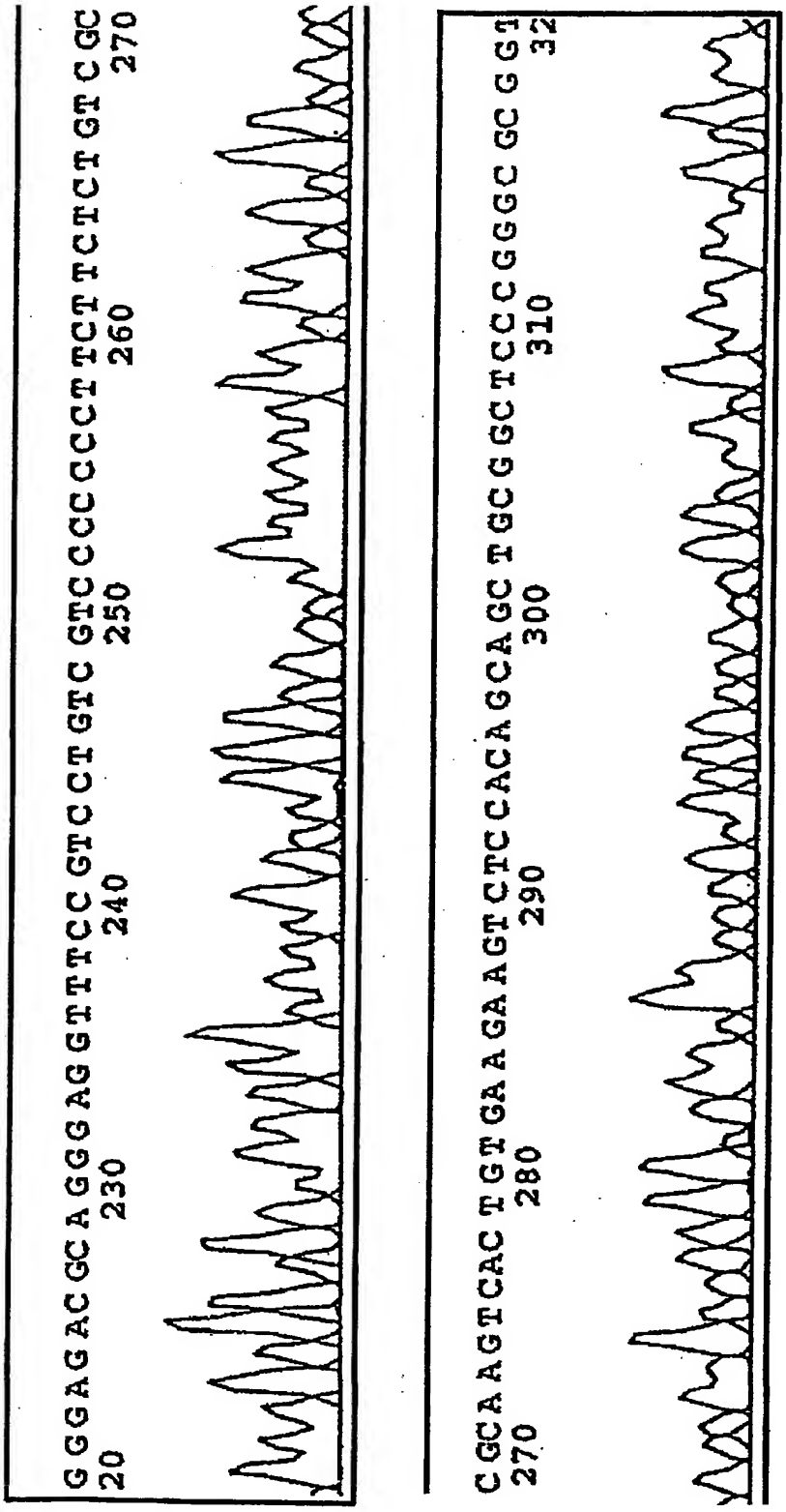

G G G A G A C G C A G G G A G G T T T C C G T C C T G T C G T C C C C C C T T C T T C T C T G T C G C
20 230 240 250 260 270
C G C A A G T C A C T G T G A A G A A G T C T C C A C A G C A G C T G C G G T C C C G G G C G C G G T
270 280 290 300 310 32

Fig.6(d)

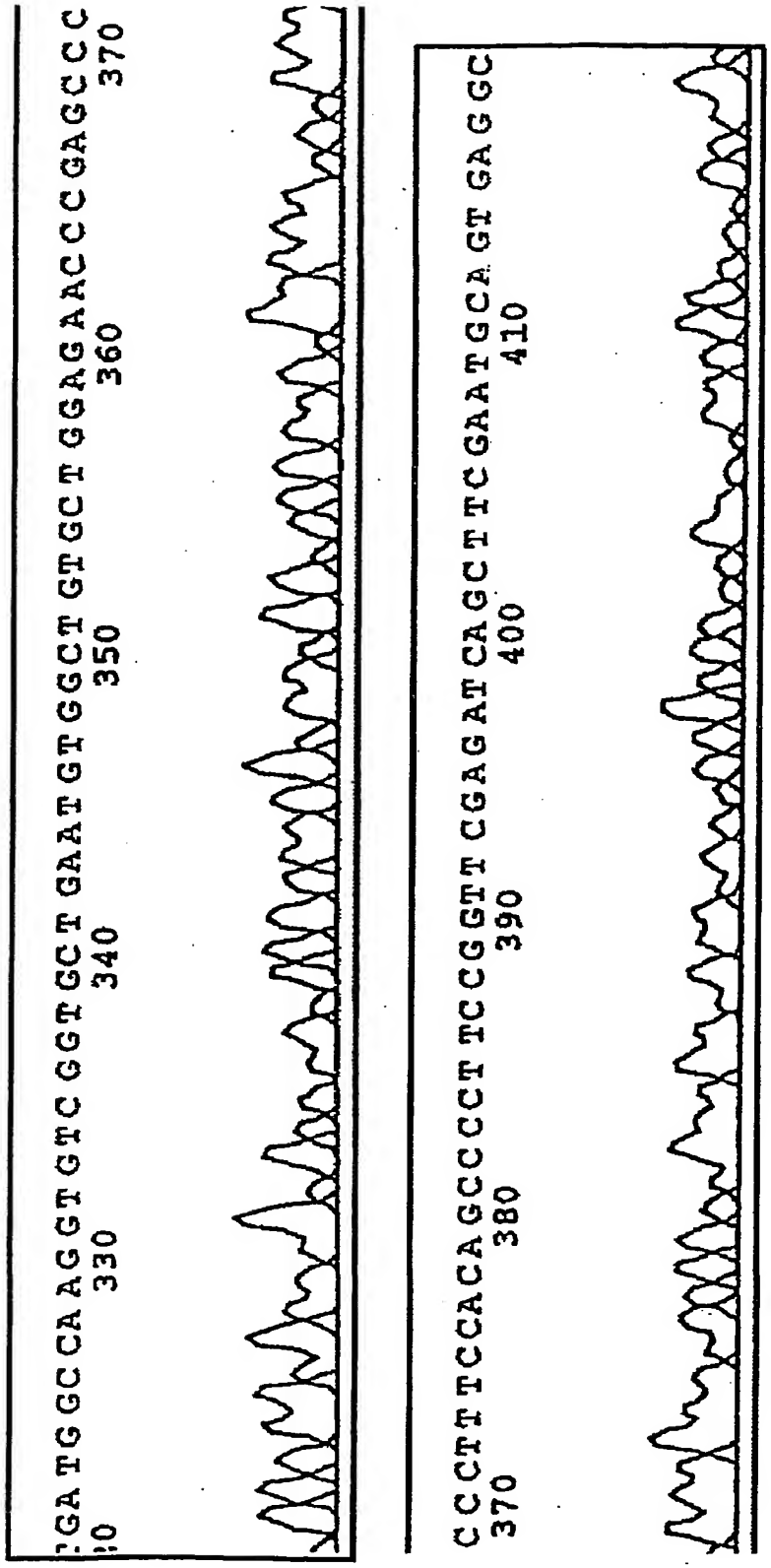

GA TG GC CA AG GT GT C GG T GC T GA AT GT GG CT GT GC T GG AG AA CC C GA GC C C
10 330 340 350 360 370
C C CT T TC CA CA GC CC CT TC C GG GT T CG AG AT CA GC T TC GA AT GC A GT GA G GC
370 380 390 400 410

Fig.6(e)

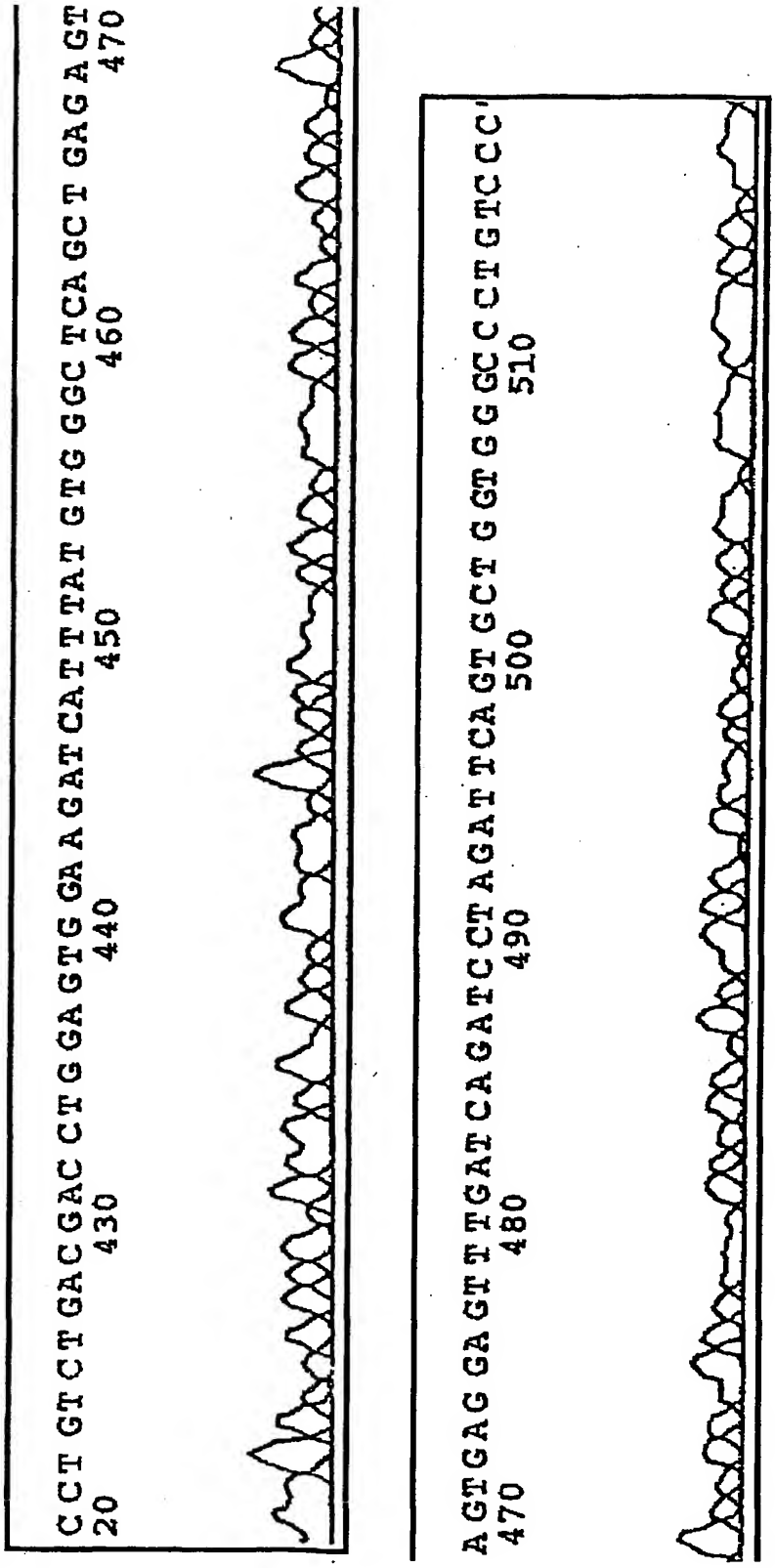

CCT GT CT GAC GAC CT G GA GTG GA A GAT CAT T TAT GTG GGC TCA GC T GA G A GT
20 430 440 450 460 470
A GTGAG GA GT T TGAT CA GA T C CT AGA T T CA GT GC T G GT G G GC C CT GTC CC'
470 480 490 500 510

Fig.7(a)

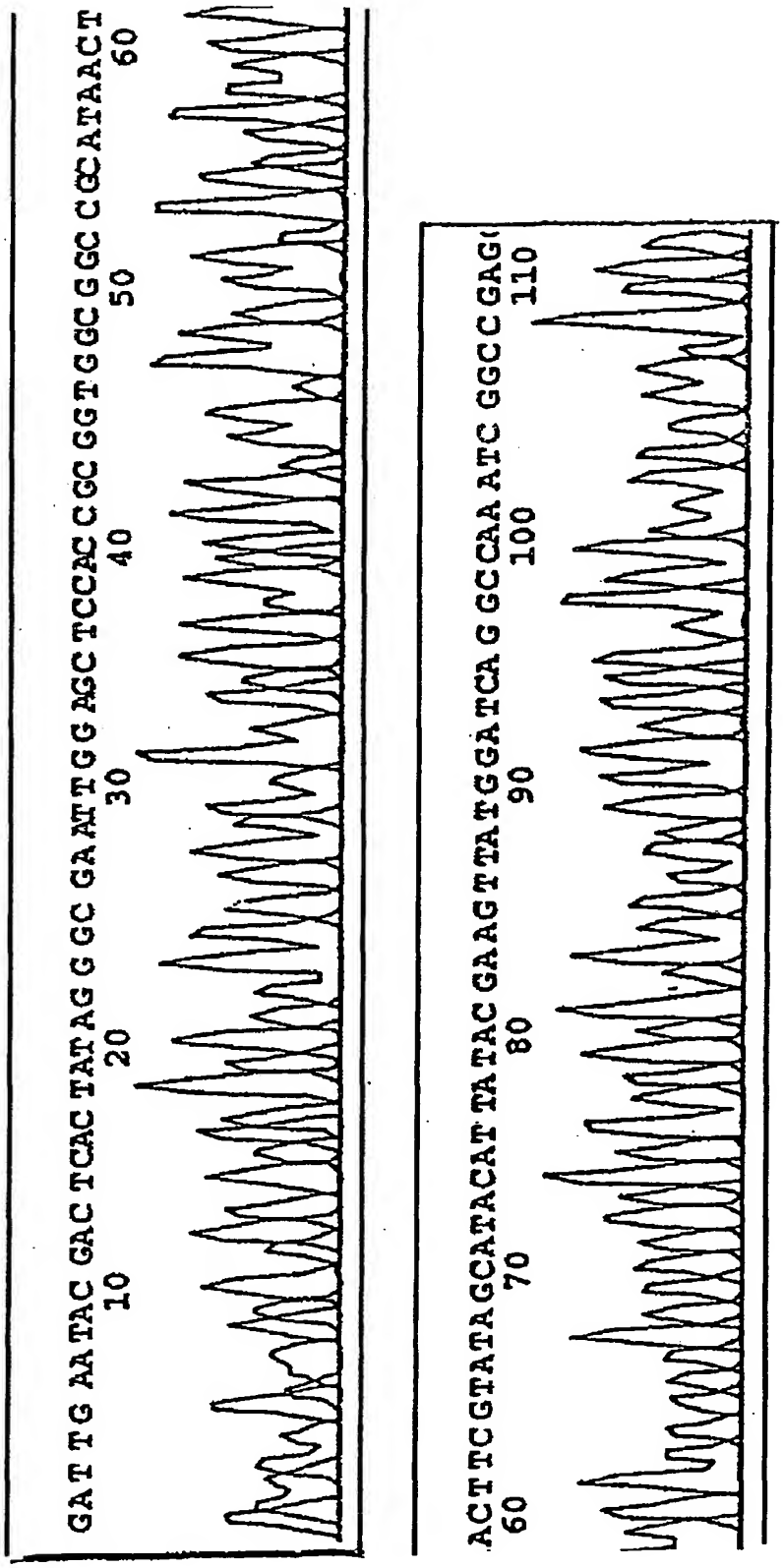

GATTG AATAC GACTCAC TATAG G GC GA ATTGG AGC TCCAC C GC GGTGGC C GC ATAACT
10 20 30 40 50 60
ACTTC GTATA GCATACAT TA TAC GA AGTTA TGGA TCA G GCCAA ATC GGCC GAG
60 70 80 90 100 110

Fig.7(b)

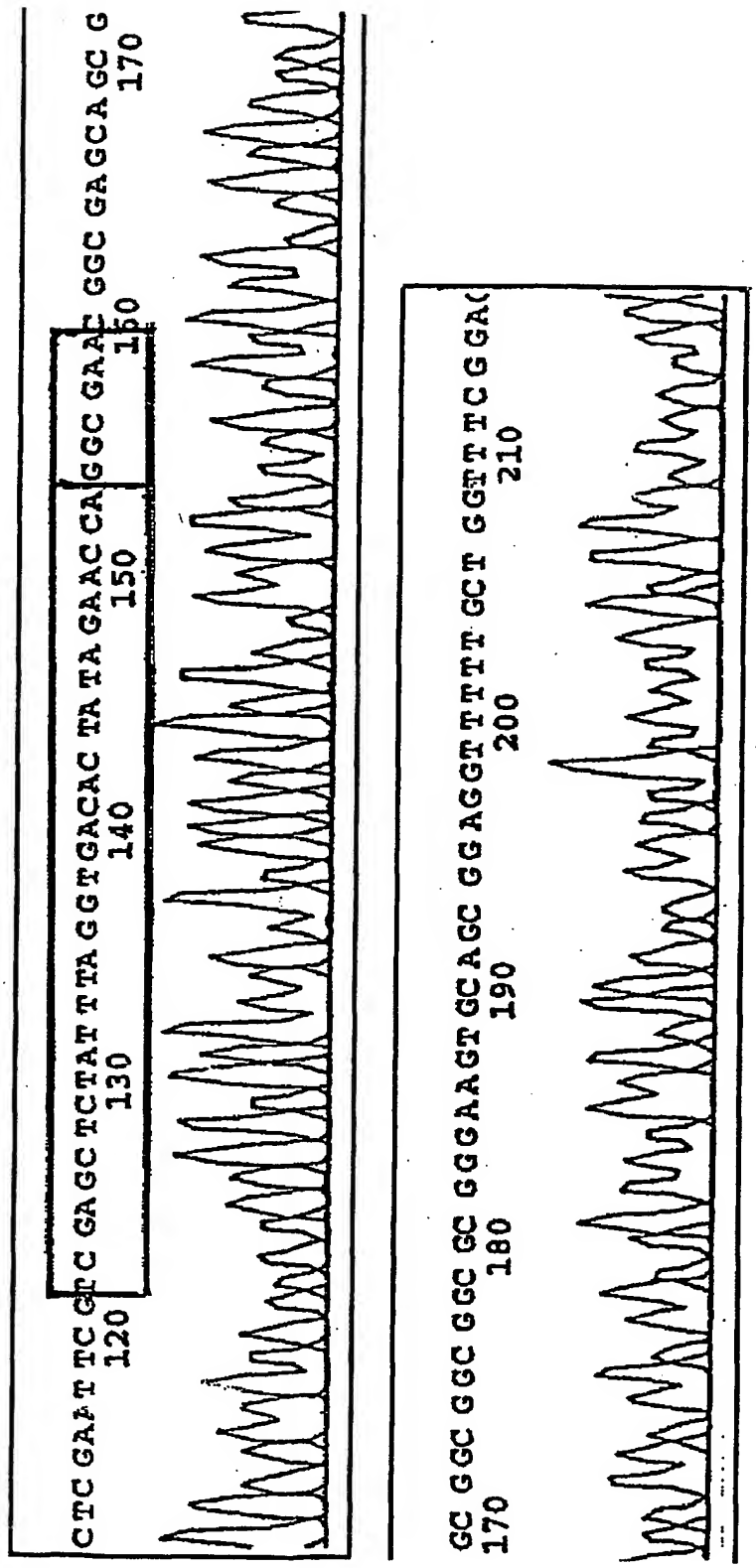

CTC GAA T TC GTC GA GC TCTAT T TA GGTGACAC TA TA GA AC CA GGC GA AC GGC GA GCA GC G
120 130 140 150 160 170
GC GGC GGC GGC GC GGGAAGT GCAGC GGAGGTTTTT GCT GGTTTCGGA
170 180 190 200 210

Fig.7(c)

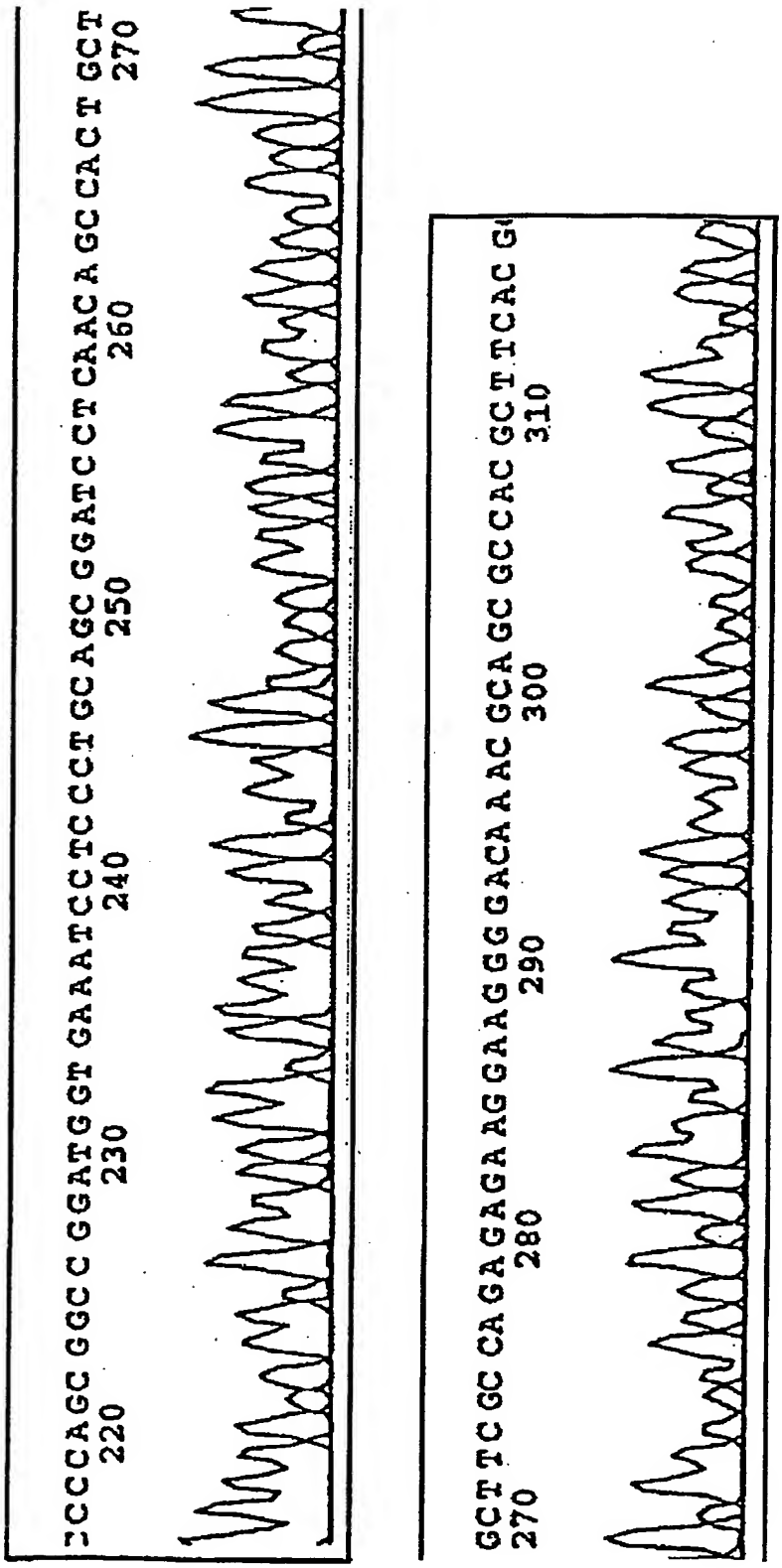

TCCCAGC GGCC GGATGGT GAAATCCTCCT CCT GC AGC GGATCCT CAAC A GC CAC T GCT
220 230 240 250 260 270
GCT TC GC CA GAGAGA AGGAAGGG GACA A AC GCA GC GC CAC GC T TCAC G
270 280 290 300 310

Fig.7(d)

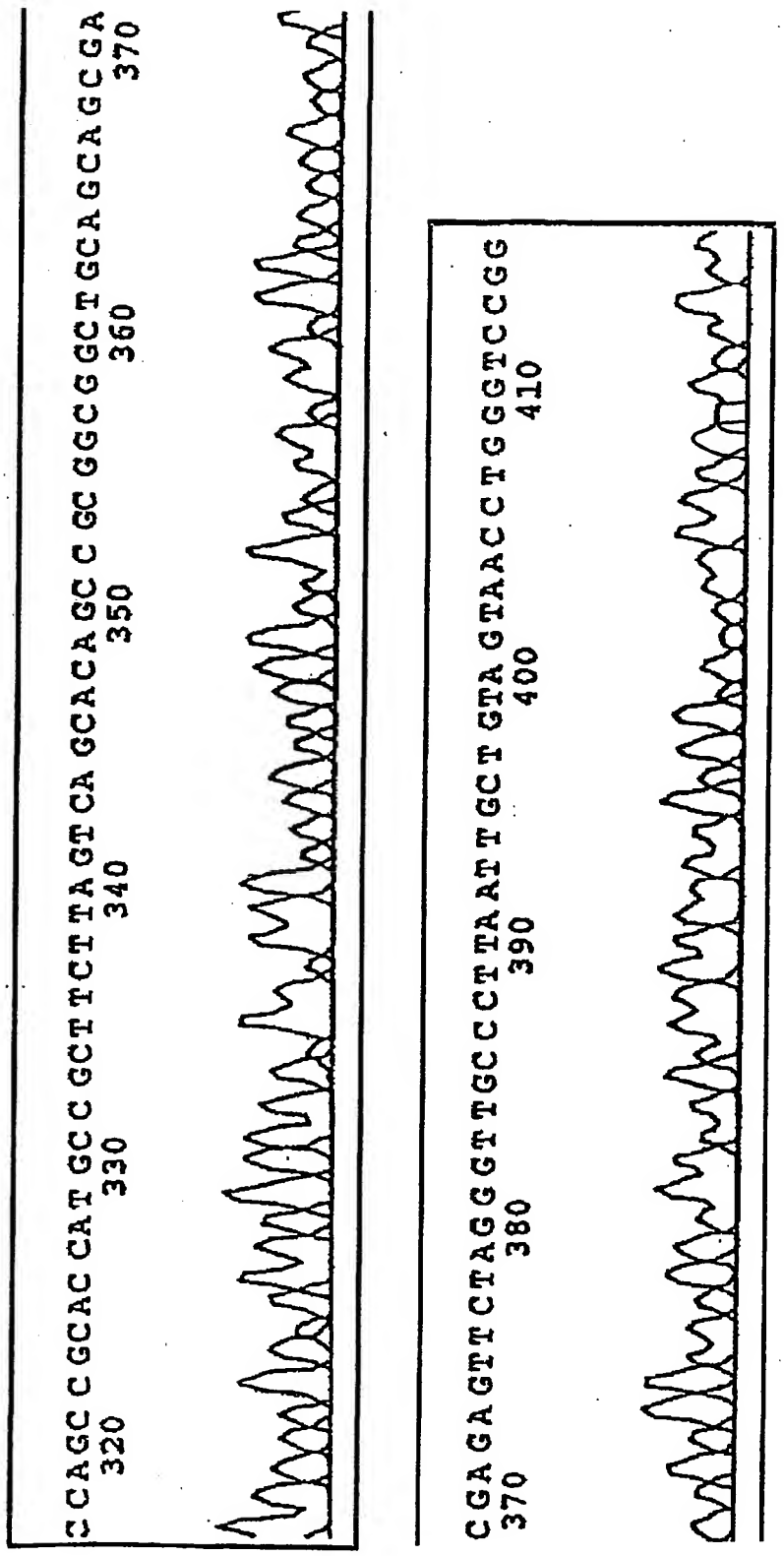

CCAGC C GCAC CA T GC C GCT T CT TA GT CA GCA CA GC C GC GGC G GC T GCA GCA GC GA
320 330 340 350 360 370
CGA GA GT T CT AGG GT T GC C CT TA AT T GC T GTA GTA AC C T G GGT CC GG
370 380 390 400 410

Fig.7(e)

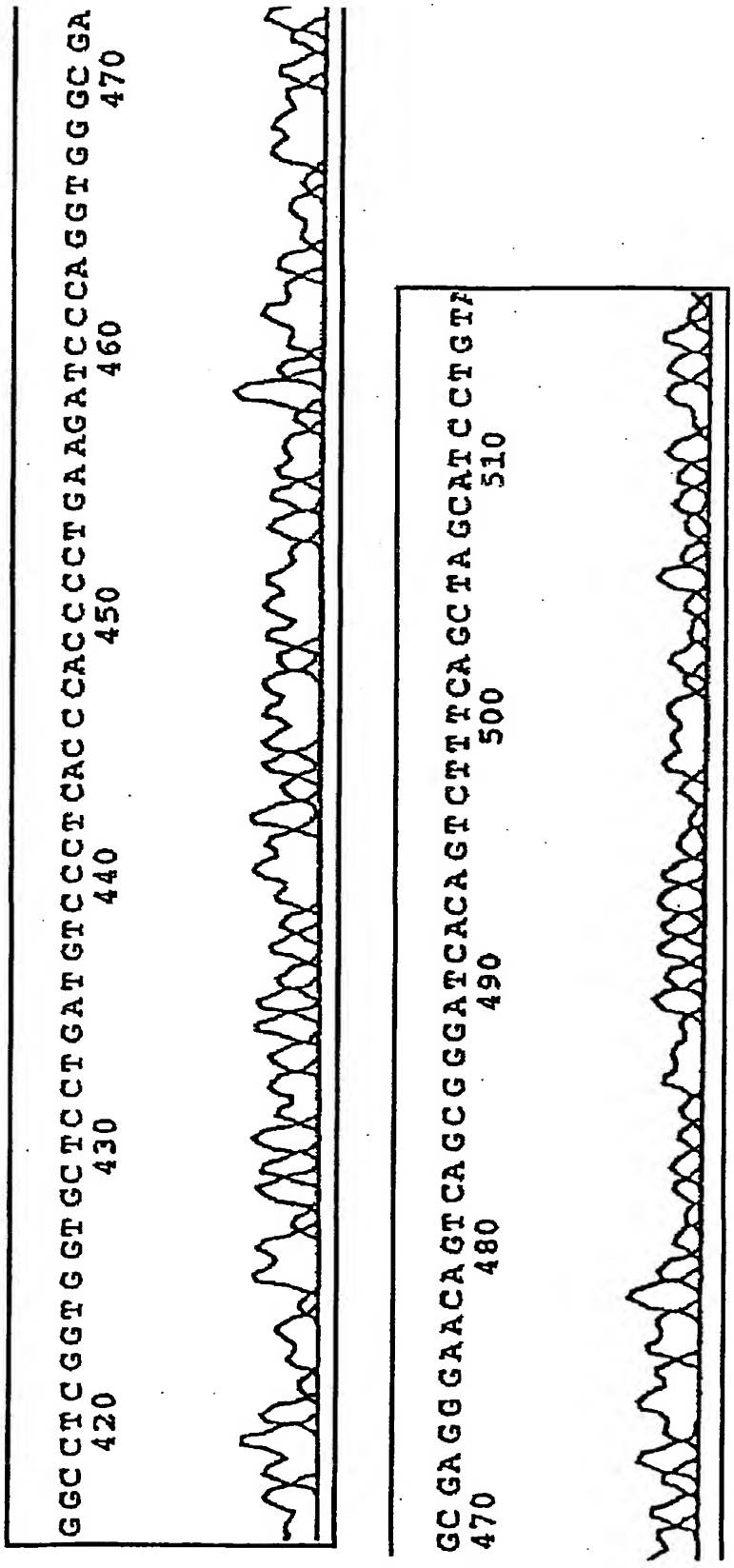
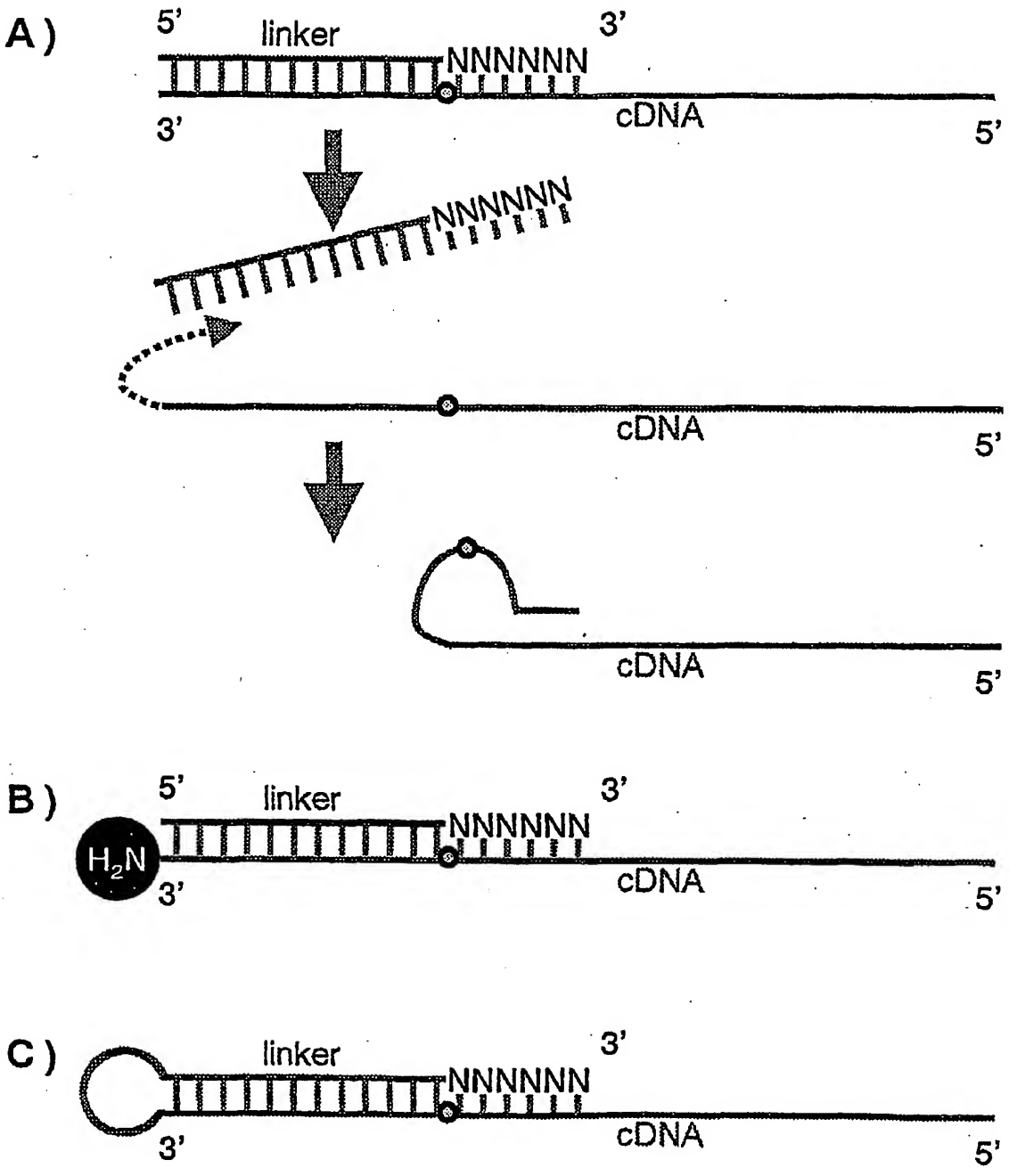

GGC CTC GGT G GT GC TCC CT GAT GTC CCT CAC C CAC C CCT GA AGA TCC CA GGT GG GC GA
420 430 440 450 460 470
GC GA GG GA ACA GT CA GC G GGA TCA CA GT CTT TCA GC TA GC AT C CT GT
470 480 490 500 510

Fig.8


A)
5'
linker
3'
NNNNNN
3'
cDNA
5'
NNNNNN
cDNA
5'
cDNA
5'
B)
5'
linker
3'
NNNNNN
H2N
3'
cDNA
5'
C)
linker
3'
NNNNNN
3'
cDNA
5'

SEQUENCE LISTING

<110> The Institute of Physical and Chemical Research

Hayashizaki, Yoshihide

<120> Oligonucleotide linkers comprising a variable cohesive portion and method for the preparation of polynucleotide libraries by using said linkers

<130> Linker Ref:1-13

<140>

<141>

<160> 7

<170> PatentIn Ver. 2.1

<210> 1

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: GN5 linker upper strand

<220>

<221> misc_feature

<222> (45-49)

<223> Nucleotides 45-49 are n wherein n = any nucleotide

<400> 1

agagagagag ctcgagctct atttaggtga cactatagaa ccagnnnnn 49

<210> 2

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: linker lower strand

<400> 2

tggttctata gtgtcaccta aatagagctc gagctctctc tct 43

<210> 3

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: N6 linker

upper strand

<220>

<221> misc_feature

<222> (44-49)

<223> Nucleotides 44-49 are n wherein n = any nucleotide

<400> 3

agagagagag ctcgagctct atttaggtga cactatagaa ccannnnnn 49

<210> 4

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer comprising BamHI site

<220>

<221> misc_feature

<222> (42)

<223> Nucleotide 42 is v wherein v = g or c or a

<220>

<221> misc_feature

<222> (43)

<223> Nucleotide 43 is n wherein n = any nucleotide

<400> 4

gagagagaga aggatccaag agctcttttt tttttttttt tvn 43

<210> 5

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: M13 forward primer

<400> 5

tgtaaaacga cggccagt 18

<210> 6

<211> 106

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: oligonucleotide

<400> 6

ggccgcataa cttcgtatag catacattat acgaagttat ggatcaggcc aaatcggccg 60

agctcgaatt cgtcgacgag agactgcagg agagaggatc cggtac 106

<210> 7

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

Oligonucleotide

<400> 7

cggatcctct ctcctgcagt ctctcgtcga cgaattcgag ctcggccgat ttggcctgat 60

ccataacttc gtataatgta tgctatacga agttatgc 98